

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 9, 2005, 13:38:21 ; Search time 30.0703 Seconds
(without alignments)
4885.990 Million cell updates/sec

Title: US-10-702-148-27
Perfect score: 7967
Sequence: 1 MAPLGRKPFPLVNPDPGEE.....NCFMMLVNTQFCALTDIVT 1527

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 65 summaries

Database : PIR 79:.*
1: piri:.*
2: piri2:.*
3: piri3:.*
4: piri4:.*

Pred. No. is the number of results predicted by chance, to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6989	87.7	1479	T17401	transcription regu
2	569	7.1	811	T08738	hypothetical prote
3	451.5	5.7	1378	G89637	protein F33HL.4 [i
4	449	5.6	1430	T34516	hypothetical prote
5	388	4.9	449	T12495	hypothetical prote
6	347.5	4.4	1264	S64146	probable membrane
7	327.5	4.1	1348	T34418	hypothetical prote
8	324	4.1	1790	S67593	transport protein
9	309.5	3.9	1526	A45605	mature-parasite-in
10	304	3.8	3225	I52300	giantin - human
11	303	3.8	3259	A45639	giantin - human
12	300.5	3.8	1372	JCS420	smooth muscle myos
13	298.5	3.7	2116	T21627	myosin heavy chain
14	295	3.7	1827	T16270	hypothetical prote
15	290.5	3.6	3187	JCS837	364K Golgi complex
16	290.5	3.6	5327	T13564	microtubule-associ
17	289.5	3.6	1938	JCS421	smooth muscle myos
18	287.5	3.6	1909	A45592	liver stage antigen
19	283.5	3.6	1979	S03166	myosin heavy chain
20	282	3.5	4687	A39638	plectin - rat
21	281.5	3.5	1372	A41604	myosin heavy chain
22	281.5	3.5	2139	T18296	myosin heavy chain
23	280.5	3.5	1390	S51364	sperm tail-specific
24	280.5	3.5	2663	S28261	centromere protein
25	280	3.5	2057	S61477	myosin II heavy ch
26	279	3.5	1819	A71928	cag island protein
27	277.5	3.5	2020	T21174	hypothetical prote
28	277	3.5	2017	A36014	myosin heavy chain
29	277	3.5	4684	A59404	plectin [imported]

ALIGNMENTS

RESULT 1

T17401

transcription regulator WBSR9 - mouse

N:Alternate names: Williams-Beuren syndrome deletion transcript 9 homolog

C:Species: Mus musculus (house mouse)

C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004

C:Accession: T17401

R:Peoples, R.J.; Cisco, M.J.; Kaplan, P.; Francke, U.

Cytogenet. Cell Genet. 82, 238-246, 1998

A:Title: Identification of the WBSR9 gene, encoding a novel transcriptional regulator,

A:Reference number: Z18735; MUID:99077764; PMID:9858827

A:Accession: T17401

A>Status: preliminary; translated from GB/EMBL/DBDJ

A:Molecule type: mRNA

A:Residues: 1-1479 <PRO>.

A:Cross-references: UNIPROT:Q92277; EMBL:AF084480; NID:g4165088; PID:g4165089; PIDN:AAD

C:Genetics:

A:Gene: Wbscr9

A:Map position: 5

F:1360-1415/Domain: bromodomain homology <BRO>

Query Match 87.7%; Score 6989; DB 2; Length 1479;

Best Local Similarity 90.3%; Pred. No. 5.9e-304;

Matches 1343; Conservative 67; Mismatches 66; Indels 12; Gaps 7;

QY 1 MAPLGRKPFPLVNPDPGEEPTTPTQEAFTREEYEARLERYSERIWTCKTGSSQL 60

|||||

Db 1 MAPLGRKPFPLVNPDPGEEPTTPTQEAFTREEYEARLERYSERIWTCKTGSSQL 60

|||||

QY 61 THKEAWESEQVAELLKSEFPNWKVLEVMVHHNTASLEKLVDSAWLIMTKYAVGEC 120

|||||

Db 61 THKEAWESEQVAELLKSEFPNWKVLEVMVHHNTASLEKLVDSAWLIMTKYAVGEC 120

|||||

cag pathogenicity
ankyrin 2, neurona
trichohyalin - hum
early endosome ant
plectin - human
trichohyalin like
hypothetical prote
nonmuscle myosin I
myosin heavy chain
probable heat shoc
myosin heavy chain
myosin heavy chain
glutamic acid-rich
repeat organellar
nestin - rat
transcription init
myosin heavy chain
myosin heavy chain
hypothetical prote
transcription acti
myosin heavy chain
nuclear/mitotic ap
nuclear migration
hypothetical prote
microtubule-vesicl
SNF2alpha protein
hypothetical coile
myosin heavy chain
myosin heavy chain
centrosome associ
NF-180 - sea lamp
myosin heavy chain
nuclear mitotic ap
hypothetical prote
helicase II - huma
myosin beta heavy


```
QY 121 DFEVGEKMLKVIKVIHPIPLEKVDREATEKKSDGACDSPSSDKENSSQIAQDHQKKEVTV 180
Db 121 DFEVGEKMLKVIKVIHPIPLEKVDREATEKKSDGACDSPSSDKENSSQIAQDHQKKEVTV 180
QY 181 KEDEGRRESINDRARRSPKLPSTSLKGERKWAPPKFLPHKYDVVKLQNEDKIISNVPADS 240
Db 181 KEDEGRRESINDRARRSPKLPSTSLKGERKWAPPKFLPHKYDVVKLQNEDKIISNVPADS 240
QY 241 LIRTERPNKEIVRYPIRHNAIRAGTGENAPWVVEDELVKKYSPLSPKFSDFLLDPYKMT 300
Db 241 LIRTERPNKEILRYPIRHNAIRAGTGENAPWVVEDELVKKYSPLSPKFSDFLLDPYKMT 300
QY 301 LNPSTKRNKTGSDPRKPKSKTDSNLSPLNPKLWCHVHLKXSLGSPPLKVNKSNK 360
Db 301 LNPSTKRNKTGSDPRKPKSKTDSNLSPLNPKLWCHVHLKXSLGSPPLKVNKSNK 360
QY 361 SPEEHLEEMKMWSP--NKLHTNFHPPKGGPPAKPGKHS DKPLKAKGRSKGILNGQKST 418
Db 361 SPEEHLEEMKMWSP--NKLHTNFHPPKGGPPAKPGKHS DKPLKAKGRSKGILNGQKST 418
QY 419 GNSKSPKGLKTPKTQMKQMTLLDMAKGTQKMTAPRNSGGTPRTSSKPHKHLPPAALHL 478
Db 420 GNSKSPKGLKTPKTQMKQMTLLDMAKGTQKMTAPRNSGGTPRTSSKPHKHLPPAALHL 478
QY 479 TAYYKENDREKRSALSCVISTARTLISSEDRARLPEELSLVQKRYELLEHKRWASM 538
Db 480 TAYYKENDREKRSALSCVISTARTLISSEDRARLPEELSLVQKRYELLEHKRWASM 538
QY 539 SEQQRKEYLKKRREELKKLKKAKERREKEMLERLEKQRYEDQELTGKNLPAFLVDT 598
Db 540 SEQQRKEYLKKRREELKKLKKAKERREKEMLERLEKQRYEDQELTGKNLPAFLVDT 598
QY 599 PEGLPNTLFGDVAMVVEFLSCYSGLLPDAQYPIITAVSLMEALSADKGGFLYLNRLVIL 658
Db 600 PEGLPNTLFGDVAMVVEFLSCYSGLLPDAQYPIITAVSLMEALSADKGGFLYLNRLVIL 658
QY 659 LQTLQDEIAEDYGBLGKLSIPIPLTHSVSELVRLCLRRSDVQESRGSDDTDNDKNSA 718
Db 660 LQTLQDEIAEDYGBLGKLSIPIPLTHSVSELVRLCLRRSDVQESRGSDDTDNDKNSA 718
QY 719 PDNEVQDEFLEKLTSEFFELTSEKLIQILFALCHRLIMTYSVDHMETROOMSAELWK 778
Db 720 PDNEVQDEFLEKLTSEFFELTSEKLIQILFALCHRLIMTYSVDHMETROOMSAELWK 778
QY 779 ERLAVLKEENDKRAEKQKREMEANKENGKVENGLKTDKRRKIVKPEPOVDTEADM 838
Db 780 ERLAVLKEENDKRAEKQKREMEANKENGKVENGLKTDKRRKIVKPEPOVDTEADM 838
QY 839 ISAVKSRRLIATQAKKERIEIQEREMKVKLERQAEERIRKHKAAAEKAFQEGIAKAKLV 898
Db 840 ISAVKSRRLIATQAKKERIEIQEREMKVKLERQAEERIRKHKAAAEKAFQEGIAKAKLV 898
QY 899 RRTPIGTDRNHNRYLWFSDEVPLGFIKGVHDSIDYRFNHHCKDHTVSGDDEYCPRSKK 958
Db 900 RRTPIGTDRNHNRYLWFSDEVPLGFIKGVHDSIDYRFNHHCKDHTVSGDDEYCPRSKK 958
QY 959 ANLGKVASMNTQGTATEVAVETTPPKQGNLWFLCDSQKELDELLNCLHPQIGRESQK 1018
Db 960 ANLGKVASMNTQGTATEVAVETTPPKQGNLWFLCDSQKELDELLNCLHPQIGRESQK 1018
QY 1019 ERLKRYQDIHSHLARKPNLGLKSCDGNQBLNPLFLRSDLEIATRLQKGLGVETTS 1078
Db 1019 ERLKRYQDIHSHLARKPNLGLKSCDGNQBLNPLFLRSDLEIATRLQKGLGVETTS 1078
QY 1079 EPEARVISLEKLDKFGECVIALQASVKKFLQGFMAKPKQRKQLOESDSAKTEEYDEKK 1138
Db 1079 EPEARVISLEKLDKFGECVIALQASVKKFLQGFMAKPKQRKQLOESDSAKTEEYDEKK 1138
QY 1139 MYEAKVASALEKWKTAIREAQTFSRMHVLLGMLDACIKWDSAEANRCKVCPKKGEDDK 1198
Db 1139 MYEAKVASALEKWKTAIREAQTFSRMHVLLGMLDACIKWDSAEANRCKVCPKKGEDDK 1198
```

```
QY 1199 LILCDECNKAFHLCRLPALYEVDPGEWQCPACQATARRNSRGRNYTEESASESDEDE 1258
Db 1199 LILCDECNKAFHLCRLPALYEVDPGEWQCPACQATARRNSRGRNYTEESASESDEDE 1258
QY 1259 S-DEEEEEEEEEEEEDYEVAGLRRLPRKTIIRGKHSVIPPAARSGRRRFGKPKPHSTRSOP 1317
Db 1259 SGRREEEEEEEEEEEEEDYEVAGLRRLPRKTIIRGKHSVIPPAARSGRRRFGKPKPHSTRSOP 1317
QY 1318 KAPPVDDDAEVDLVLQTKRSSRRQSLQKCEELHKKIVKRYFSPFPFVTRDEADYY 1377
Db 1318 K----DDPEVDLVLQTKRISRRQSLQKCEELHKKIVKRYFSPFPFVTRDEADYY 1377
QY 1378 DVIHPMDPOTVQNKSCSGSVRSVOEFLTDKMQVFTNAEVNCRGSHVLSQWVKTEQCILV 1437
Db 1378 DVIHPMDPOTVQNKSCSGSVRSVOEFLTDKMQVFTNAEVNCRGSHVLSQWVKTEQCILV 1437
QY 1438 VLLKHLPGHPYVRRKRKFFDRLAEDGDSSEPAVGQSRDEDRRSR 1485
Db 1438 VLLKHLPGHPYVRRKRKFFDRLAEDGDSSEPAVGQSRDEDRRSR 1485
QY 1439 VLLKHLPGHPYVRRKRKFFDRLAEDGDSSEPAVGQSRDEDRRSR 1485
Db 1439 VLLKHLPGHPYVRRKRKFFDRLAEDGDSSEPAVGQSRDEDRRSR 1485
RESULT 2
T08738
hypochemical protein DKFp586E0518.1 - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 09-Jul-2004
R:Wambutt, R.; Heubner, D.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, March 1999
A:Reference number: Z16471
A:Accession: T08738
A:Molecule type: mRNA
A:Residues: 1-811 <WAM>
A:Cross-references: UNIPROT:Q9NRL2; EMBL:AL050089
A:Experimental source: adult uterus; clone DKFp586E0518
C:Genetics:
A:Note: DKFp586E0518.1
F:709-764/Domain: bromodomain homology <BRO>
Query Match 7.1%; Score 569; DB 2; Length 811;
Best Local Similarity 24.4%; Pred. No. 3.6e-18;
Matches 184; Conservative 115; Mismatches 239; Indels 216; Gaps 24;
QY 872 BEERIRKHKAAAEKAFQEGIAKAKLVMERITPIGTDRNHNRYLWFSDEVPLGFIK 928
Db 29 EEEALKQHQREKELLEKIQSALACTNIFPLGRDMYRRYWF-PSIPGLFIEDYSG 87
QY 929 VHDISI----DYRFNHHCKDHTVSGDEDYCPRSKKN-LGKNASMNTQGTATEVAVETT 982
Db 88 TEDMLLPSPSSFQNNVQSDPQVS-----TKTGEPLMSESTNIDOG-PRDHSVQLP 138
QY 983 TPQGNLWFLCDSQKELDELLNCLHPQIGRESQKLERL--EK-----RYQDIHSHI 1032
Db 139 KPVHKPNRWCYSSCEQLDLTEALNSRGHRESALKETLLOEKSRICQLARFSE--EKF 196
QY 1033 HLARKPNLGLK-----SCDGNQ-----ELLNPLRSDLEIATRLQKGLGVETT 1077
Db 197 HPSDKPQDPSKPTYSRGRSSNAYDPSQCAEKQLELRDLFDIEDRIYQGTGLGAKVT 256
QY 1078 -----SFEFA---RVISLEKL-----KDFGECVIALQASVKKFLQGFMAKPKQR 1119
Db 257 DRHWRSALESGRYELLSEENKENGIIKTVDNEVEEMEIDEQTKIVK--DRLLGKLTET 314
QY 1120 RKLQSEDSAKTEEVDE-----EKQVBEAKVAS-----ALEKWKTA 1155
Db 315 PSTVSTNASTPQSVSSVVHYLAMALFQIEQGIERRFLKAPLDASDSGRSYKTVLDWRRES 374
QY 1156 IREACTFSRMHVLLGMLDACIKWDSAEANRCKVCPKKGEDDKLILCDECNKAFHLCRL 1215
Db 375 LLSSASLSQVFLHSLTDRSVIWSKSIILNARCKICRKKGDAENMVLDCGDRGHHTYCVR 434
QY 1216 PALYEVDPGEWQCPACQATARRNSRGRNYTEESASESDESDDEDEEBE-----EBE 1271
Db 1216 PALYEVDPGEWQCPACQATARRNSRGRNYTEESASESDESDDEDEEBE-----EBE 1271
```

Db 435 PKLKTVPEDGFCPCRCQKRSRRLSSRRQPSLESDEDEDVDSMGDEDEVDGDEEBSQSE 494
QY 1272 EDEYEV-----AGLRRLPRKTI-----RGKHSVIPPAARSGRRPGKXPHSTR 1313
Db 495 EEEYEVEQDEDDSQEEBEEVSLPKRGPQVRLPVKTRGKLSSSFSRGGQEPGRYPSRSQ 554
QY 1314 RSOPK-----APVD-----DAEVEL-----1330
Db 555 QSTPKTTVSKTGRSLRKINSAPPTTKSLRIASRSTRSHSGHLOADVFFVLLSPRRKR 614
QY 1331 -----VLOTKRSSRQSL-----1343
Db 615 GRKSANKTPNSPNFNPFRVIATKSSEQSRVNIASKLSQESKRXCKKRSQSPSPV 674
QY 1344 -----ELQKCEILHKIVKYRFSFPFVTRDEADYDVITHPMDFTV 1389
Db 675 TLGRRSGRGGVHLSAFEQLVVELVRHDDSWFLKLSKIQVPDYDIKPIALNII 734
QY 1390 QNKCSGYSRSVQEFILTMKQVFTNAEVNCRGS 1423
Db 735 REKVNKCEYKLASEFIDDIELMFNSCFEYNPRNT 768

RESULT 3
G88637
protein F53H1.4 [imported] - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 09-Jul-2004
C:Accession: G88637
R:Anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biology
A:Reference number: A75000; PMID:99069613; PMID:9851916
A>Note: see websites genome.wustl.edu/gsc/C_elegans/ and www.sanger.ac.uk/Projects/C_elegans/
A>Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
A:Accession: G88637
A>Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-1378 <STO>
A:Cross-References: UNIPROT:O45075; GB:chr_IV; PIDN:AAC02578.1; PID:G2854159; GSPDB:GN00
A:Gene: F53H1.4
A:Map position: 4

Query Match 5.7%; Score 451.5; DB 2; Length 1378;
Best Local Similarity 19.8%; Pred. No. 1.2e-12;
Matches 317; Conservative 195; Mismatches 531; Indels 555; Gaps 59;

QY 71 EVAELLKEFPAPWEKLVLEWHHTASLEKLVDTAWLIMTKYAVGECDFEVGKERML 130
Db 74 EIEDLIDEXFFAG-EKLMSGKEYTVVSSSEKRGGLTLYTMEDGKTIGHR---DLRRKGL 129
QY 131 KYKIVK---IHPLEKYDEATEKSDGCDSP-----SSDKENSSQIAQ-----171
Db 130 VEEIKKIAIEAEFVDEKQVRELLAENPIREIKKYAPFSANRKSPTPKAQLSVAA 189
QY 172 --DHQKKTVVKDEGRRESINDRARRSPKPLTSLKGERKWAPPKFLPHYDVKLQNE 229
Db 190 EADSDVQVSDMAESGAANKSAMKTPRGAP-----RASGGPVILSSR---RLQEK 239
QY 230 DKIIISNPADSLIRTPPNKBEIVRYFIRHNALRAGTGGENAPWVEDELVKYSLPSKFS 289
Db 240 QKEKDEKEKELEKQKEQ-----BEKKQKKEEAKK--LKEK-E 277
QY 290 DFLDPPKYWTNLPSTKRNKTSQPKSKSKTNSLSSPLNPKLWCHVHLKLSGS 349
Db 278 EKLKEEKAARAEKKEKNNGTMDK-----FLKXD-TGS 311
QY 350 PLKVNSKN-----SKSPEEHEEMKQWSPNKLHFNPHIPKGGPAPKPGKSHSDKPLK 403
Db 312 P-----SSKNAPLFSFSKWEKRIAVGVKQ-----EDAKRRDLELY 349
QY 404 AKGRS----KGILNGQKSTGNSKSPKGLKTPKTKMKQMTLLDMAKGTQKWTAPRNSGCT 460

Db 350 NEACSWCEKNLSGNORST-----FENPIFKFSVQKLVDRKAK-----385
QY 461 PRTSSKPHKHLPAALHLIAYVYKENKDRKDSALSCVISTARLLSSSDRRLPEELRS 520
Db 386 -----HMKGKWAQKAE--FKAEMSEKREKELYQKPEPKI KAFWNEDI-----428
QY 521 LVQKRYELLEHKK--RWASMSBEQRKEYLKKRBEELKKLKKAKERREKEMLERLEKQK 578
Db 389 -----HMKGKWAQKAE--FKAEMSEKREKELYQKPEPKI KAFWNEDI-----428
QY 579 RYEDQELTKNLPAP--RLVDTPEGLPNTLFGDVAMVVEFLSCYSGLLLPDAQYPITAVS 636
Db 429 ALDDLVTCTENLELLDAKRVNCDDELLKCL-----EISQFFVSMRKILLWNEN--ITAEQ 481
QY 637 LMEALSADGGFLYLN-RVLVILLQTLQDEIAEDYGEKLMKLSSEPLTLHSHVSELVRLC 695
Db 482 LRDDLHGGDFGPRSTYKMIANLLETALQEKHEKAAHCNARLSEFPINEHTTISELIRAF 541
QY 696 LRRS-----DVQBESEGS-----TDDNKDSAAAFEDNE 723
Db 542 FIGSTETSPKRDGKRAAGAHDDDDDEEDGDEEMDSREKVVETEKEKPEENVGAANGE 601
QY 724 VODEFLEKLETSEFFELTSEKQLQILTALCHRLIMTYSVODHMETROQKSAELWKE-----779
Db 602 IGDDEEENSES-----ELQKQRIALFADSCHIIYELPAGAQ--LEVLCA-----KEVVG 650
QY 780 ---RLAVLKEENDKKRAEKQK---RKEMAKNKE-----NGKVENGKLGKTDK---821
Db 651 LPIIREWFLURDANSEKLTETHEKCANRINEMEQHOQLQDFPPIPEITESTMTQTRAE 710
QY 822 ---KRIVKPEQVDTAEADMSIAVKSRRLLAIQAKKERIOBREMKVKLERQABEERIK 878
Db 711 QSLRRREKLENQDIT-----LKIELENREATAREVD-DLER-----746
QY 879 HKAAAEKAPQEGIAKAKLVMRTPITGTRNHRNYLFP---SDEVP-----GLPIEKGVWH 930
Db 747 -----IFRVVHIGNDRHLRKYWFAYSDAAIWWQDFGTTSYSEKVV 788
QY 931 DSIDYRFNHCKDHTVSGDEDYCPRSKANLGNASMTQHGATATEVAVETTPKQGNL 990
Db 789 DCEKGF-----DVESSDVENRPEYEDLPL-----TSSQSET 822
QY 991 WFLCDSQKELDELLNCLHPQGIRESOLKERLEKRYQDIHSHLARKPNLGLKSCDGNQ 1050
Db 823 WYKLDTEPAIRQLMTQKNGKREKLLKYLNNMDDIISSILRKEKQ---KKDGEER 879
QY 1051 -----LLNF 1054
Db 880 DBEASEDEDEASAAENGEBEEMETEONGTGKFGFPETAAAAAAAEQBAENARRFTGRFGS 939
QY 1055 LRSDLIEVATRLQKGLGVVEETSFEARVVISLEKLKDFG-----ECVI---A 1099
Db 940 LKRTSELNLDWKQSGISKIVDSQVFPEARMLEBANTLDEMRLVTELVTIPVECVIEKFP 999
QY 1100 LOASVIKKFLQGFMAPKQKRRKLQSED-----SAKT-----EEVDEEKQW 1140
Db 1000 ONVAIAKFLNFTSTILLEKKNYEFSDPFRACRIPFNYPFLFSHKTSKLEKLNQTMKIG 1059
QY 1141 BEAKVASALEKWKTA-----IREAQTFSRMHVLLGLMDLACIKWD 1179
Db 1060 AKFGICAIQKPKDVRLLPKKKECFSLKMPFCRRVQEQASNASCLHMLLAYFDARIDQ 1119
QY 1180 MSAENARCKVCPKGEDDKLILDCSNKAFHLFCLRP-----ALYE---VPDGEWQCPAQ 1232
Db 1120 RTLPELSQVCHRRKTGTERKLMCKOCSTVFHYGCHRPTRISRALFEEGPKEG-WWCAKCT 1178
QY 1233 PATAARNRGRNYTBESASESED-----DESDEEEEEEEEEDEVEAGLR 1283
Db 1179 KEDRR-----QLSEAKEDLRQEEGGDEEDHHGSGESDDEDEDIVVEETRGR 1228
QY 1284 PKRTIRKGHSVIPPAARSGRRPGKXPHSTRSQKAPP-----VDDAEVDELVLQTKRS 1337

Db 1229 SAKR-----KANAAMRDVLEFEGVLROTAPPKPKQKTVVVPVLELNFLSIERA 1278
 Qy 1338 SRQSLQKCEILHKIVKYPSPFPFVRTRDAEDYDVITHPMDPOTVQNKSCGS 1397
 Db 1279 NPLRYKTLQOI-----PGQSRSTRNAQHNSRLPDIQDLDV-----1315
 Qy 1398 YRSVOBELTDMQVFTNA---EVVNCRGSHVLSVMVK 1431
 Db 1316 YTSALQHLHLSQFFRHARGYIETHNPRKFSKSPFK 1353

RESULT 4
 T34516
 hypothetical protein ZK783.4 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 09-Jul-2004
 C:Accession: T34516
 R:Favella, A.; Vaudin, M.
 submitted to the EMBL Data Library, August 1994
 A:Description: The sequence of C. elegans cosmid ZK783.
 A:Reference number: Z21536
 A:Accession: T34516
 A>Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1430 <FAV>
 A:Cross-references: UNIPROT:Q23590; EMBL:U13646; PIDN:AAC24421.1; GSPDB:GN00021; CESP:ZK
 A:Experimental source: strain Bristol N2; clone ZK783
 C:Genetics:
 A:Gene: CESP:ZK783.4
 A:Map position: 3
 A:Introns: 248/3; 373/3; 547/1; 593/2; 905/2; 1042/3; 1116/3; 1317/2; 1376/1
 F:1284-1339/Domain: bromodomain homology <BRO>

Query Match 5.6%; Score 449; DB 2; Length 1430;
 Best Local Similarity 18.7%; Pred. No. 1.6e-12;
 Matches 305; Conservative 257; Mismatches 576; Indels 494; Gaps 62;

Qy 126 KEKMLKVIKVIHLEKVEATEKSDGACDS-----PSSDKENSQ-----168
 Db 21 QOOLQOQLAKI---QKATSPSKSTNGTSASTSAVSTSGTSSQNEAAQLQNLAKWQ 77
 Qy 169 -----IAQDHQKKEITV-----KEDEGRRESINDRARRSPKLPSTSLKKGK 211
 Db 78 QIQQLAQFALMAAQKQKEKAADKAKEKEKQKAAAAAASASSTSSAIPG 137
 Qy 212 WAPPKPLPHYDVQKQNEDKIISNVPADSLITERPPNKEIVRYFIRHNALRAGTG 271
 Db 138 LSPMLAAWQQAQIQM-----ALQQMMWTPQKSQM-----167
 Qy 272 WVVEDELVKYSLPSKFSDFLLDPKYMTLN-PSYKRNKTSFDRKPSKSKTD-----324
 Db 168 -----BEAIKK-----MMDMAKKKPGAVASTSSASTSTPSTSSASITSSNNNA 213
 Qy 325 NSGLSPLNPKLVCHVHLK-----KSLSGSPKVKNSKNSKSP 362
 Db 214 NNAASNNWNVWQLVAAQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQ 272
 Qy 363 EBHEBEMMK-MMSPNKLTNPHIPKPGPPKPKGSHDKPLKAKGRSKGILNGKSTGNS 421
 Db 273 ----BEILFLMAQHGLN-----HKKGHEKQADAALAAKVAHRAALES 315
 Qy 422 KSPKGLKTPKTKMQLTLDMAKGTQKMT-----RAPNSGGTPTSS 465
 Db 316 DSPEGKKTNEAMLR-----LPIQLGWRRQTCVRSIASAGVGDVSYFAP--CGKKLSTYS 369
 Qy 466 KPHKLPAPALHLIAYKENKREDKRSALSCVSKTARLLSSEDR---ARLPEELRSIV 522
 Db 370 EVVRVLTNKSIIHYIT--RNFLENYKLVIGFIVPKQTEADTEQEREFAMTEDDINKE 427
 Qy 523 QKRYELLEH-KKRWASMSBEQKEYLKKKR-----BELKKLKE-----560
 Db 428 LTRNLVLPKVPKIQASTNGVHDDIKMSKIBEPDPLDPSELNDEFTBELVHSQIMSNG 487

Qy 561 ----KAKERREKEMLERLEKQKRYEDQBLTGKLPAPRLVDTPTEGLPNTLFGDVAVVVF 616
 Db 488 VDECKIRREADDLLVNDV-----RHLPDFSRIGN-QCLSSQGFADALMVHEF 536
 Qy 617 LSCYSGLLPDAQYPITAVLSMEALSADKGFLYLNRVLLVILLQTLQDEIAEDGELGM 676
 Db 537 VQNFHVLGIDLEIAPKLESACAGLDGDAN--HAEQTLQLTRQLL--RLALEPFGMG 590
 Qy 677 KL-----SEIPLTHSVSELVRLCLRRSDVQEESESGSDTDNDKSDAAAFEDNEVQDEFLE 730
 Db 591 EKRFQGGGEMGLDRENFSEVNLFL-----DKGKRG-----EELSQ 628
 Qy 731 KLETSEFFELTSEEKLOILTALCHRLMTYSV-----QDHMETROQMSAELW-----777
 Db 629 PLLTCNFLISPEQKASILAFLCDELVCSRVVTEIDKNLDELISRLKGEKWNREGKARAL 688
 Qy 778 -----KERLAVLKEE-NDKKRAEKQKKE-----MEAKNKENGK 810
 Db 689 RSARSKKNDEKVVVVKEEQNHESDEPPTPDKATVAPPTVVSVSPVSAQAQQRK 748
 Qy 811 VENGGLKTRKRIYKVPFQVDTAEADMISAVKSRRLAIQAKEREIQEREMKVKLRQ 870
 Db 749 FTPGLGQC-----EVLTEQESMSLQQMSDLIGDLHQEAQINQK-----788
 Qy 871 AEEERIRKHAATAEAKAFOEGIAKALVMRRTPIGTDNRNRYWLF--SDEVPGLFIEKW 928
 Db 789 -----IHDTCIKIRSPFGTDRFRNRYWMLAHTDKV---IIESL 824
 Qy 929 VHDSDYRHNHCKDHTVSGDEDCPRSKKANLGNASMTQHGTPATEVAVETT-----982
 Db 825 ATTSVN--NPACNANEVASKDP--PTLEQRPVGCETIDLDVIACTEDLVDDVLLRAK 879
 Qy 983 ----TPKQGNL-----WFLCDSQKELDELNCLHPQIGRESOLKERLEKRYODIHS 1031
 Db 880 ADKTRKRYRIENHMKRGWMTQNRDCVESLRSCLSGIRERARHLRLTKPW-----933
 Qy 1032 IHLARKPNLGLKSCDGNQELNLFRLSDIEVATRLQKGLGVVEETSFEARVISLEK 1091
 Db 934 -----FLNELKFGITTEPVGEKSDLELVRQGWTRLNT-AIDKL- 972
 Qy 1092 DFGCVIALQASVIKKFLQGP-----MAPKQKRKQSEDSA-----KTEBV 1133
 Db 973 ---QC--HLKMSDVSKPLSPITPPETQKPIVVVPTMALAQIVKDDMAWKVIDEEDVGQBL 1027
 Qy 1134 DE---EKWVEAKV---ASALEKWKTAI-REACTFSEMHVLLGLMCLACIWDKSAENA 1185
 Db 1028 DETTIQKIIETADMDQDTQCLFEDMKSYVSTEAOQTSQLMVALQTLGEMINWERSREA 1087
 Qy 1186 RCKVCPKGGEDDKLILDCBNKAFHLCRLPALVEVPDGEWOCQACQATARRN-----1239
 Db 1088 LCQIC-KSMDGDEMLVCDGCSGCHMECFRPMTKVPEGDWFCQRCREKSGRPMCMFCS 1146
 Qy 1240 -----SRGRNYTESASDESDESDDEE-----EEEE 1268
 Db 1147 RETGNLHQQRCAVHVHQCSDQGPKEAINPETFCGHQCEMKQMFVKRLLILRSESESR 1206
 Qy 1269 EEEEDYEVAGLURLPRKTIIRKGVIPPAARSG---RRPGKPHSTRSQKAPVDDA 1325
 Db 1207 ELEDNDHAENG-----ENTQNGHNMNGAIGVHNQNGVYKGNLKRK-----L 1251
 Qy 1326 EYDELVLQTKSRRRQSLQKCEILHKIVKRYFSPFPFVRTRDAEDYDVITHPMD 1385
 Db 1252 EVPS-----GGLPKNNKELCQMLDELVVQANALFLEPVNKLVPYKMIISKPD 1305
 Qy 1386 FQTVQNKCSGYSRQVBEFLTDMQVFTNAEVYVNCRGSHVLSVMVKTEBQCLVLLHKLHP 1445
 Db 1306 LKTIQKNEKLIYETPEDFAEDIELMFANCRQFNIDHSEIGRAGIS-----LHKFPQ 1357
 Qy 1446 -----GHPYVRKR-----KFP-----DRL---AEDEGDSF-----PEAVG 1474
 Db 1358 KWKQLKYNFTKRLRLHPKQYPIHLRHTCSTLSERLVAFVAVAGFAEQARTLGAAG 1417
 Qy 1475 QSRDEDRSREA 1486

Db 1418 SSEELEAPSTVA 1429

RESULT 5

T12495

hypothetical protein DKFZp434H071.1 - human (fragment)

C:Species: Homo sapiens (man)

C>Date: 23-Jul-1999 #sequence_revision 23-Jul-1999 #text_change 09-Jul-2004

C:Accession: T12495

R:Poustka, A.; Klein, M.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.

submitted to the Protein Sequence Database, June 1999

A:Reference number: Z117525

A:Accession: T12495

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-449 <POU>

A:Cross-references: UNIPROT:Q9UIF8; EMBL:AL080173

A:Experimental source: adult testis; clone DKFZp434H071

C:Genetics:

A:Note: DKFZp434H071.1

C:Superfamily: transcription factor GCN5; bromodomain homology

F:366-421/Domain: bromodomain homology <BRO>

Query Match 4.9%; Score 388; DB 2; Length 449;

Best Local Similarity 23.3%; Pred. No. 2.1e-10;

Matches 111; Conservative 81; Mismatches 169; Indels 116; Gaps 12;

QY 998 KEDELNCLHPOGIRESOEKERLEKRYODIHSIHLAKPNLGLKSCDGNQELNLFARS 1057

Db 6 EDLKALLKVLHRLGIREKALQIQK-HLDYITQACLKNK-DVAIIELNENEE--NQVTR 61

QY .1058 DLIE---VATRLQKGLGVYEETSEPEARVISLEKLGKDFECVIALQASVIRKFLQGFMA 1114

Db 62 DIVENWSVEEQAMENDSLVQLQVEDLERVAS-----ASLQ-----VKGWNC 103

QY 1115 PKQ-----KRRKQSEDSAKTEEVDEBK 1137

Db 104 PEPASEREDLVYFEHKSFTKLCKEHDGEFTGEDSSAHLERKSONPLDIAVTRLADLER 163

QY 1138 KMVE--EAKVASALEKWKTAIREAQTFSRMHVLLGMLDACIKWDSAEARNCVKCPKGE 1195

Db 164 NIERIEEDAPGLRVRRLSEARSAAQVLCIOQLKSIKAWKSIEMKVCQICRKGDN 223

QY 1196 DDKLILCDENKAFHFLCLRPALYVEPDGEWCPCAPATA-----RRNSRGRNYTEE 1248

Db 224 EELLLLCDGCKGCHTYCHRPKITIPDGMFCPACIAKASGQTLKIKLHVKGKKNES 283

QY 1249 SASE-----DSEDDSEDEEEEEEEEEEEEDYEVAGLRPRKTRIRGHSHVPPAARG 1302

Db 284 KKGKVVTLTGDTEDSDS-----ASTSSS 306

QY 1303 RPPGKKPHSTRSQKAPPVDDAEVDVLQTKRSRRSRLQKCEILHLKIVKRFWSW 1362

Db 307 LKRGNKDLKRRKMEENTS--INLSKQESFTSVKPKRDSKDLALCSMLTTEMETHEDAW 364

QY 1363 PPREPTRDEADYDVITHPDMDFQVONKSCGSYSRVSQVEFLTDMKQVFTNAEYVN 1419

Db 365 PLLLPNKLVPGYKKVKKPKMDFSTIREKLSGGQYPNLETALDVLVDFDNCETFN 421

RESULT 6

S64146

probable membrane protein VGL13w - Yeast (Saccharomyces cerevisiae)

N:Alternate names: hypothetical protein G2842

C:Species: Saccharomyces cerevisiae

C>Date: 17-May-1996 #sequence_revision 17-May-1996 #text_change 09-Jul-2004

C:Accession: S64146; S71739

R:Escibano, V.; Eraso, P.; Portillo, F.; Mazon, M.J.

submitted to the Protein Sequence Database, May 1996

A:Reference number: S64144

A:Accession: S64146

A:Molecule type: DNA

A:Residues: 1-1264 <ESC>

A:Cross-references: UNIPROT:P53125; EMBL:Z72655; NID:g1322701; PID:e243361; PID:g132270

A:Experimental source: strain S288C

R:Escibano, V.; Eraso, P.; Portillo, F.; Mazon, M.J.

Yeast 12, 887-892, 1996

A:Title: Sequence analysis of a 14.6 kb DNA fragment of Saccharomyces cerevisiae chromo

A:Reference number: S71733; MUID:96437978; PMID:8840506

A:Accession: S71739

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-1264 <ESW>

A:Cross-references: EMBL:X92670

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1995

C:Genetics:

A:Gene: SGD:ITC1

A:Cross-references: SGD:S0003101

A:Map position: 7L

C:Keywords: transmembrane protein

F:549-565/Domain: transmembrane #status predicted <TMM>

Query Match 4.4%; Score 347.5; DB 2; Length 1264;

Best Local Similarity 19.5%; Pred. No. 4.5e-08;

Matches 271; Conservative 206; Mismatches 492; Indels 419; Gaps 58;

QY 3 PLLGRKPPFLVNPPLGPEEPFFTIPTQEAFTREEYEARELERYSERIWTCKSTGSSQLTH 62

Db 8 PILLPDPKPL--PLDLNVQVWHIEETGEWFSSVEEFLEFDFVTRHHFTCEITGTSLTLP 65

QY 63 KEAMEEQEVALLKEBPAPWYKLVLEMVHIN-TASLEKLVDTAWLEIMTKYAVGEED 121

Db 66 FOALDSEETQFYKVEDRFPFLKLRPEVARFLHFNGIRRLDALVEKYVARFNDPFPGE--- 122

QY 122 FEVGKEMLVKVIKVIHPLEKVDDEATEKKSGACDPSDDKENS--QTAQDHQKE 177

Db 123 -----VYLKQKDSSTSSNSQOSTPQDDMVEINSVGNPGLPYQYQRR 168

QY 178 TVYKDEDEGRRESINDRARRSPKLPSTLKKGRKWAPPFLPHKYDVQKLVQNEDKIISN-- 235

Db 169 VYIKVKQVFNATINPESREIVNPAHT-----KY---MLIEEAASNSKS 208

QY 236 --VPADSLIRTRPPNKELVYFIRHNALRAGTENAPWVVEDELVKYKSLSKFSDFLL 293

Db 209 FIVDQGIYRDRSTFTKHLIKCFKITLQRASSKMGAPCVKPEYLAAMYGLTMEWPK--- 265

QY 294 DPYKYMTLNPSTKRK---NTGSPD-----RKPSKSKTDNSLSPLNPKLWCHVHL 342

Db 266 DMLKYEDEPVVARNSGANVSSPESEKVKROSKSGKSNSTNDASNKKETK--- 318

QY 343 KKSLSGSPKLVKNSKNKSPSEHLEBEMMKWSPNKLHTNFHPKKGPPAKPKGKHSKPL 402

Db 319 ----XKKPTEVNDSENNSEED-----KKGQNVTS 347

QY 403 KAKGRSGILNQKSTGNSKSPKGLKTPTKQKQMTLL-----DMAKQTQKMTAPRNS 457

Db 348 HSKRKKKE-ANBEPNTENVES---VPTPANAEPQAVTITSIMDLA----- 389

QY 458 GGTPTSSKPKHGLPPAALHLIAYKENKDRDKRSALSCVISKARLLSSSDRLRLEE 517

Db 390 -----LPYOH-PFNIPFNITYNEK-----LECI-----SLGSTKLSRPFDS 425

QY 518 LRSVLQKRYELLE---HKRWASMSSEQRKEYLK-----KKREEL 554

Db 426 FGKLLQ-AVQFLNTFGSKICLSHFSLDQITSKCTDPVELKGEVVLVNIROTSTKEQEI 484

QY 555 KKK---LKEAKERKEKEMLERLEKQRYEDELTKG---NLPAFLV-DTPEG-----LPN 604

Db 485 ENNGLPKMKNAFTTTEEDSENPSDQWNSFIRDMIMKRNDSKVEYKIVHDDPASDILN 544

QY 605 TLFGDVAMVVEFLSCYSGLLLPDAQYPTAVSLMEALSADKGGFLVNLVILLOTLLQ 664

Db 545 INHNGSALLIEVFTALLRFI-NBERGDSGCVIENWIIDKG-----VLMER--K 591

```

QY 665 DEIAEDYDGLGKMLSEIPLTLHVSSELVRL--CLRR--SDVQSEEG-SDTDDNKDSAAF 719
DQ 592 DERGE--GEAKQKRAHGYFLOQKEKIDNLDKTLKENATEVQKESDAKNETNSDSKSD 649
QY 720 EDNEVDQDEFLEK-----LETSEFFELTSE----- 743
DQ 650 SDSEERDPKLEKLYNRVNMWIERLTQKFNNSYWLIIILGLVLEDCRHLPMWTEFIDSFI 709
QY 744 EKL-----QIITLALCHRI-----LMTY-----SVQDHMETRO 770
DQ 710 EKIIIPKDISATQPKQLWRNFCRLKLSFSDKVNALMILVDLVSHPDIAKAAVDDSM--- 766
QY 771 QMSAELMWERLAVLEENDKKAQKQKEMEA-----KKNKGVKVENGLKTDKRRKI 824
DQ 767 -LCGQIRSERFKVARELKTAAVLSNLOGLDQAIQEKLNKTDENTPSADGADKKD----- 820
QY 825 VKFEPQVTEADMTSAVKSRLLAIAQAKEREIQEREMKVKLRQABEERIRKHAASAE 884
DQ 821 ---DSENSPEIDLIIEKKQKLEEQDKVQALQ-----SD 854
QY 885 KAPOGIAKAKLVMRTPIGTDNRNRY--WLPSEVDP-----GLPIKGMWHDIDY-- 935
DQ 855 KNFLDNCLFENDLQRLKPLGLDRYGNRYFWLDHNGVPPQYPAGM--NETPKSNNSLSYHS 913
QY 936 -----RNHCKDH-----TVSGDEDYCPRS 956
DQ 914 GRLLIQGKASSAKFVLNVDSEQLSNWQKIRNSEGISARETVFGISKTSGSYNVENG 973
QY 957 KKAN--LGRNASMNTQ--HGTATVAVETTPKQ--GONLWFLCDQKELDELNCLHPQG 1011
DQ 974 IVEVLDSNDRVNPILIELTPICKIMDETSPSRLLSPOWYICDKLEDLSRIMDWLNDWG 1033
QY 1012 IRESOLKERLEKRYQDIITHSIHLAKPNLGLKSCDGNQELL-----NFRSLDI 1060
DQ 1034 RKEHLLAQIRPIMERIKSSLSL--RDHALSLTAFTKNEEKLKLENNENFTENELNVDSM 1092
QY 1061 EVATRLQGLGVYETSEFEARVLSLEKLDKDFGECVIALQASVIKFLQGFMAPQKRR 1120
DQ 1093 DVDDK--NSGV-----KSEVDVQVDAEEK-----REAVIDEKLEVIADLE 1130
QY 1121 KLOSDSDAKTEEV-----DEEKMVEE-----AKVASALEKMKTAIREAQ 1161
DQ 1131 -MKLDDSKTRVNLNRIQELDQORDELLEQKSIINSQRPGARILARSERKTKISRGNK 1189
QY 1162 FSRMVL 1169
DQ 1190 VNQIEIL 1197

RESULT 7
T34418
hypothetical protein F12F3.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T34418
R:Fulton, B.; Wohldmann, P.
submitted to the EMBL Data Library, July 1998
A:Description: The sequence of C. elegans cosmid F12F3.
A:Reference number: 221521
C:Accession: T34418
A:Status: Preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-3488 <FUL>
A:Cross-References: EMBL:U80022; PIDN:AAC5885.1; GSPDB:GN00023; CESP:F12F3.3
A:Experimental source: strain Bristol N2; clone F12F3
C:Genetics:
A:Gene: CESP:F12F3.3
A:Map position: 5
A:Insertions: 281/3; 332/1; 562/3; 600/3; 1866/3; 1944/3; 3393/1

Query Match 4.1%; Score 327.5; DB 2; Length 3488;
Best Local Similarity 19.1%; Pred. No. 1.1e-06;
Matches 334; Conservative 230; Mismatches 619; Indels 563; Gaps 75;

```

```

QY 36 EYEYARLERYSERIW-----TCKTSGSSQLTHKEAWEEQVAVELLLKBEFPDAP--YEKLV 89
DQ 280 DRMEVRHDEDEWKRKWLILKIDCKD-----EEAEYACQAINVAGEAWCSDVVV 327
QY 90 EMVH--NTASLEKLVDTAMLEIMTKYAVGECDFEVGEKMKLVK-----IVKIHP 139
DQ 328 HNSEESRDDKSVDEVDSTVLE--EKDGDGDKSKPKTKKI IKKETPESEQVTAAP 385
QY 140 ----LEKYDEBATEKSDGACDSSPKSQAQDHQKXETVVYKEDGRRSINDRAR 195
DQ 386 EQOKISEVDVQSAETEVAEKKPAEKPTDLSAKDKSKG--KSDEPEASTEESTT 442
QY 196 RSPKPLPSLKKGERKWAAPPK--FLPHKYDVKLQNEDKIISNVPADSLIRTERPP----- 248
DQ 443 EKPTNDKTSKSAEKKTVKPKKEVTGKPLEAKKPVEDKKDASQPSS--KSSPPTDGKK 500
QY 249 NKEIVRYFIRNALRAGTGENAPWVEDELVKYKSLSPKFSDF----- 291
DQ 501 KKQIPKALF-----IPDEISRFGDPSTMHSETNITTTIRREGSADAK 544
QY 292 --LLDP-----YKMTLNPSTK-----RKNTGSPDRK----- 316
DQ 545 TPLVEPLSASVMKVFITLVEAKAEFSFKRRSETPDCKRKEGLPPAKKSEKDEVT 604
QY 317 -----PSKSKTDNSSL--SPLNPKLWCHVHLKLSLSSPLKVKNSKNSPPE 364
DQ 605 AEKOSTEALIESKKEVDKISQOQSDKN-----KSEVVGVPEKAAGPETKDVSE 657
QY 365 HLEEMKMSPNKLTNPHIPKGGPPAKPGKHSDKPLKAKG----- 406
DQ 658 -LEE-----VPKKTKIKKTEK--SDSSISQKSNVLKPADDKSKSDVD 700
QY 407 RSKGLNGOKSTGNSKSPKGLKTPKTKMKQWTLTD-----MAKGTQKMTAPRNSGG 459
DQ 701 KSKKTTEDQTKVATDSKLEKAADTK-QIETETVDDSKKKVLLKTEKSDPSFSQSE 759
QY 460 TPRSSSKPHKLPALHLLIAYYKENDRED-----KRSA-----LSCVIS 500
DQ 760 TPPV--VEPTKPAESEAQIAEVNKAQKQEVDDNLKREAVAAKIADEKLKIEANIK 818
QY 501 KTRALLSEDRARLPEELR--SLVQKR--YELLEHKRWASMSBEQKRYLKKRE--EL 554
DQ 819 KTAVEAAKQKQKDEQKLETEVVSKSAEKLKLEKQ--AQIKKAAEADAVKKQKELNE 877
QY 555 KKKLKEKAKERREKEMLER-----LEKQRYEDQELTGKULPAFLVDTTPGLPNT 605
DQ 878 KNLKAAKSAADKLKLEESAASKSVSESVKFGEEKTKAGEKTVQVESEPTSKTI 937
QY 606 LFGDVAMVVEFLSCYSGLLLPDAQYPIITAVSLMEALSADKGGFLYNRVILLQTLQD 665
DQ 938 DTKDV-----GATEPADETPKKIKIKKTEKSDSS-----ISQ 970
QY 666 EIAEDYDGLGKMLSEIPLTLHVSSELVRLCLRRSDVQSESGSDTDDNKDSAAFNEVQ 725
DQ 971 KSATDSEKVSQKQDEPTKPAVSE-----TQMTVTEADSKKQKQKDEKLKDAEIA 1022
QY 726 DEFLEKLETSFFELTSEKIQILALCHRIIMTVSVQDHMETROQMAELMKE--RLAVL 784
DQ 1023 AK--TKQEADEKSKLDAQEKIK-----KVSDEDAARKKEKELNDKLKLESEIATK 1069
QY 785 KEENDKKAQKQ-----KKEKEMAKNKENGKVENGLKTDKRRKIRVKEFPQVDTEADMTS 840
DQ 1070 KASADKLKLEBOAQAKAAEVAACKQKQEKDE-----QKLTDEAASKKA 1114
QY 841 AVKSRRL-----AIOAKKEREIQEREM-----KVKLRQABEERIRKHK 880
DQ 1115 AAEKLEKQKQAKIKAAGADAVKKQKELDEKNKLEANKSAAGKLKIE--EESAASKQ 1171
QY 881 AAAEKAFQEGIAKAKLVMRTPIGTDNRNRYWLPSEVDPGLFIKGMWHDIDIRFNH 940
DQ 1172 TVEQAKLDAQTKAKTAETKQTKLEKD-----EKS----- 1200

```



```
QY 1090 LKDFGECVIALQASVKKFLQGFMAPKQKRRKLQSDSAKTEEVDEEKKMVEEAKV-ASA 1148
DQ 2163 LQDDRRDVIDEAKWKEFSDAIOQKEBEIRLKEDNCVSLKDQLRQMSIHMEELKINISR 2222
QY 1149 LEK-----WKTAREAQTSRM-----HVL-----LGMIDAC 1175
DQ 2223 LEHDKQIWES---KAQTEVQIQKVCVDITLQGENKELLSQLBETHLYHSSQNELAKLESE 2279
QY 1176 IK-----WMSAENARCKVCPKGGEDDKLLCDECNKAFHLCLRPALPYEYDGEWQCP 1229
DQ 2280 LKSLKDLQTLDSLSLEKCK--EQKGNLGGIIRQEQAD-----IQNSKFSVEQLETDLQ 2330
QY 1230 ACQATARNRGRNYTEE-----SASEDS-----EDDESE-----EEEE 1266
DQ 2331 ASRELTSLHEE- INMKQKIISLGSKEEAIOVAIELRQOHDKKEIKELENLLSQSEEE 2389
QY 1267 E---EEEEEDEVAGLRLRPKRTIRGKHSVTPPAARSGRRPGKPHSTRSQKAPVD 1323
DQ 2390 NIVLEENKKAVDKTNQLMETLTKIK-KENI-----QQK----- 2422
QY 1324 DAEVDDELV-----LQTKRS---SRRQSLQKCEBILHKIVKRFSPFRPVTTRDEA-- 1373
DQ 2423 -AQLDSFVKSMSSQLNDRDRIVGDYQLEERHLSIILEK-----DQLIQEAAAE 2470
QY 1374 -----EDYDVITHPMDQTQVQKSCGYSYRSVQEFLLDMQKQVFNVAEYVNCGRSHVLSC 1428
DQ 2471 NNLKEEIRGLRSHMDLNSNAKLDA---ELIQYREDLQNVIT----- 2511
QY 1429 MVKTEQCLVLLKHLPCHYPVRRKXKFPDRLAEDGDSRPEAVGQSRDSDRRSREAEI 1488
DQ 2512 -IKDSQ-QKQLEVLQONKELENKYAKLEBKLESE-----EANDLRRSFNA-L 2559
QY 1489 QEWLQDTS--LYSAXIN 1503
DQ 2560 QEEKQDLSKEITESLKV 2576

RESULT 11
A:56539
N:Altitin - human
C:Species: Homo sapiens (man)
C:Date: 19-Oct-1995 #sequence_revision 26-Jan-1996 #text_change 09-Jul-2004
C:Accession: A56539; S37536
R:Seelig, H.P.; Schranz, P.; Schroeter, H.; Wiemann, C.; Griffiths, G.; Renz, M.
Mol. Cell. Biol. 14, 2564-2576, 1994
A:Title: Molecular genetic analyses of a 376-kilodalton Golgi complex membrane protein
A:Reference number: A56539; MUID:94187728; PMID:7511208
A:Accession: A56539
A:Molecule type: mRNA
A:Residues: 1-3259 <SEE>
A:Cross-references: UNIPROT:Q14789; EMBL:X75304; NID:G405714; PIDN:CAA53052.1; PID:G4057
C:Genetics:
A:Gene: GDB:GOLGB1; GCP: GCP371
A:Cross-references: GDB:454958
A:Map position: 3q13.31-3q13.31
C:Superfamily: Giantin
C:Keywords: coiled coil; Golgi apparatus; transmembrane protein
F:3238-3254/Domain: transmembrane #status predicted <TM>

Query Match 3.8%; Score 303; DB 1; Length 3259;
Best Local Similarity 19.3%; Pred. No. 1.3e-05;
Matches 348; Conservative 285; Mismatches 614; Indels 560; Gaps 85;

QY 30 EAFRTREYERLERYSRITCKSTG-----SSOLTHKEAWEQEVAELLKERFPA-- 82
DQ 1031 ETERGEVEDKENKYSKCVTSKQCEIYVLKQIISKEV--ELQHIRKOLEEKLAAEE 1088
QY 83 WYEKLIVLEK-----VHNTASLEKLV-----DTAWLEIMTKYAV 116
DQ 1089 QFQALVKQMNTLQDKTQIDLLQAEISENQAIQKLIITSNTSDSGDSVAL-VKETVVI 1147
```

```
QY 117 GEECDPEVG-----KERMLKVIKVIHPLEKVDDEATEKKS--DGACDSPSSDKN 165
DQ 1148 SPPC---TGSSEHWKPELEEKILALEKEKQLOKQLQALTSRKAILKKAQKEKHLREE 1204
QY 166 SSQAQDHQKKEIVVKEDEGRESINDRARRSPKRLPTSLKKGERKWAAPPKPLPHKYDV 225
DQ 1205 LKQKDDYNNLQEQDEQSKENENIGDQLRQIQVRESI-----DQK 1247
QY 226 LONED--KIINVPA-----DSLIRTERPPNKEIVRYFIRHNALRAG 265
DQ 1248 LPSTDOQESCSSTFGLERPLFKATEQHHTQVPLESNLCPDWPSHSEDA-----SALQGG 1301
QY 266 TG-ENAPVWVDELVKYSLPSKESDFLLDPYKMTLNPSTKRKNTGSPDRKPSKSKTD 324
DQ 1302 TSVAIQKALKEIBAEKVELELVSS-----TSELTKKSEVFLQEQINKQGLE 1352
QY 325 NSSLSPPNPKLWCHVH---LKKSLSGSPKLVKNSKNSKSPPEHLEENMMKMSPNK--- 377
DQ 1353 IESLKTYSHE--AEVHAESLQQLLESSQLQIAGLEHLRELQPKLDELQKLKSKKEEDVS 1409
QY 378 -----LHTNPHIPKGGPPAKPGGHSK----- 400
DQ 1410 YLSGQSEKEAALTKIOTEIIEQEDLIKALHTQLEMQAK-----BHERIKOLOVEL 1461
QY 401 -PLKAGRSKILNQKQSTGNSK-----SPKGLKTPKTKMOMTLLDMAGTK--QKMT 452
DQ 1462 CEMQKQPEIGESRAKQOIORKLOALISKEALKENKLSQELSL---ANGTTERLTK 1518
QY 453 APRNSGGTPTTSSPKHKLPPAALHLIAYYKENDR-----EDRSALSQSVISKT 502
DQ 1519 SLADVESQVSAQNK-----KDTVLGRALLQEQERDKLITEMDRSILENQSLSSCESLKL 1574
QY 503 ARLJSSDRALPPELBSL-----VOKRYELL-----EKKRW- 536
DQ 1575 ALEGITDEKELVKEIESLKSSTIAESTEWQKHKELOKYEIILQSYENVSNEAERIOH 1634
QY 537 ---SNSEORKEYLKKKEELKKLKKAKERRKEMLERLEKOKRY----- 580
DQ 1635 VEAVERQEKLYGLKLRSTANKKETEQLOQAQEMEMKEKMKRFAKSKQOKILELEE 1694
QY 581 EDQELTGKMLPAFLVDTPEGLPNTLFGDVAMVVEFLSCYSGLLLPDAQYFITAVALMEA 640
DQ 1695 ENDRLEAEVHA-----GDTAK-----ECMETLLSSNA-----SMKEE 1727
QY 641 LSADKGGFLYNNRVILVILQTLQDEIADYVGEKMLSEIPLTHSVSELVRLCLRRSD 700
DQ 1728 LERVQMEYETLSKXQSLMSE--KDSLSEVQDLKQIETD-----NVSK----- 1769
QY 701 VQESSEGGSDTDNDKDSAAFEDNE-VQDEFLEKLETSBEFFELTSBEKLQILTALCHRLMT 759
DQ 1770 -QANLEATEKHNDQNTVTEGTQSIPTGETEQDLSMSTRPTCESVPSAKSANPAVKD 1828
QY 760 YSVODHMETROQMGAELWKERLAVLKEENDKKR-----AEQKQ----- 797
DQ 1829 FSSHDEINNYLQIDQL--KERIAGLEEKQKNEPSQTLNEKNLTLSQLSTKQELKWL 1887
QY 798 RKEMAKKNGKGVNGLGKTRDKRIVKFPQVDVTEADMIKSAVSKRLL----- 848
DQ 1888 QEEVTKMNLNQIQEEL-----SRVTKLKETAEEKDDL-----BERLMQALNGLS 1936
QY 849 -----AIQAKKEREIOEREM---KVKLEROAEEERIRKHAAA----KAFQEGIAK 893
DQ 1937 IGVYQDVTDQIKNELLESEMKNLKKCVSELEEKQOLVKEKTVSEIRKEYLEKIQG 1996
QY 894 AKLVMRRTPIGTDNRNHRNYLFSDEVPGLFIE-----KGWVHDSIDYRNFHHKCHDHTVSG 948
DQ 1997 A-----QKEP--GNKSH-----AKELQELLKEQEVKQLOKDCIRYQEKISALERTVKA 2044
QY 949 DEDYCPRSKK--ANLGKVASMNTQHG-----TATEVAVETTTKQOQNLWFLCDSQKE 999
DQ 2045 LEFVQTESQKOLEITIKENLAQAVEHRKKAQELASFKVLDDT---QSEARVADNLKL 2101
QY 1000 LDELLNCLHPQCIRE---SOLKER---LEKRYQDIHSHILARKPNLGLKSCDGNQOELIN 1053
```

Db 2102 KREL-----QSNKESVSKQMKQKQEDLERRUEQ-AEEKHLKEKQNM-----QEKLD 2146
Qy 1054 FLRSLDI-----LVATRLQGG--LGYVEE-----TSEFARVTSLEKJDFGECVIA 1099
Db 2147 ALRREKVLHEETIGETQVTLNKKQVQQLQENLDSTVQLAAFTKSMSSLODDRDRVID 2206
Qy 1100 LQASVTKKFLQFMAPKQKRLQSDSAKTBEVDEEKQWBEAKV-ASALEK-----WKT 1154
Db 2207 EAKWKEKFSDAIQSKEEBEIRLKEDNCVLDKQRMQSIHMBELKINISRLHEHDXQWES 2266
Qy 1155 AIREAQTFGRM-----HVL-----LGMLDACIK-----WD 1179
Db 2267 -----KATEVQLQKVCDDTLOGENKELLSOLETRHLYHSSQNELAKLESELKSLDKQLTD 2323
Qy 1180 MSAENARCVCPKGGDDXLLILCDECNKAFHLCRLPALYEPVDPGEWQCPACQATARRN 1239
Db 2324 LNSLEKCK--BQKGNLEGIIRQOEAD-----IQNSKFSYEQLTDLQASRELTSLRH 2374
Qy 1240 SGRNVTET-----SASDS-----EDDESDE-----EEEEEE 1273
Db 2375 EE-INMKEQKIISLSSGKEEAQVATAELRQOHDKBEIKELNLLSQEEBENIVLBEENKK 2433
Qy 1274 DYEAGLRRLPRKTIRGKHSVIPPAPARSRRPKKPHSTRSQPKAPPVDDAEVDLV-- 1331
Db 2434 AVDTKNQMETLTKIK-KENI-----QOK-----AQDSFVKS 2465
Qy 1332 ---LQTKRS---SRQSLQKCEELIKVIRKFSWPPREPVTDEA-----EDYVD 1378
Db 2466 MSSLQNDRRDRIVDYQQLBERHLSILEK-----DQLQEAANENKLEERIG 2514
Qy 1379 VITHPMDFTQVONKSCGSYRSVQBEFLTMKQVFTNAEVNCRGSHVLSVMYKTEQCLVV 1438
Db 2515 LRSHMDDLNSNAKLDA-----ELIQVREDLNQVIT-----IKDSQ-QKQ 2553
Qy 1439 LLHKHLPCHPYRRKKEKPPDLADEGDSPEAVQCSDEDRRSREARIEQWLODTS-- 1496
Db 2554 LLEVQLQONKLENKYAKLEELKSE-----EANEDLRRSFNA-LQBEKQDLSKE 2603
Qy 1497 LYSAKIN 1503
Db 2604 IESLKVS 2610
RESULT 12
JC5420
smooth muscle myosin heavy chain 1 - mouse
C:Species: Mus musculus (house mouse)
C>Date: 10-Jun-1997 #sequence_revision 18-Jul-1997 #text_change 09-Jul-2004
C:Accession: JC5420
R:Hasegawa, K.; Arakawa, E.; Oda, S.; Matsuda, Y.
Biochem. Biophys. Res. Commun. 232, 313-316, 1997
A:Title: Molecular cloning and expression of murine smooth muscle myosin heavy chains.
A:Reference number: JC5420; PMID:97242182; PMID:9125171
A:Accession: JC5420
A:Molecule type: mRNA
A:Residues: 1-1972 <HAS>
A:Cross-references: UNIPROT:008638; DBJ:D85923; NID:g1945077; PIDN:BAAL19690.1; PID:g194
A:Experimental source: smooth muscle
C:Comment: this protein plays a role in smooth muscle cell contraction.
C:Superfamily: myosin heavy chain; myosin motor domain homology
C:Keywords: nucleotide binding; P-loop
F:88-771/Domain: myosin motor domain homology <MMOT>
F:178-185/Region: nucleotide-binding motif A (P-loop)
Query Match 3.8%; Score 300.5; DB 2; Length 1972;
Best Local Similarity 19.4%; Pred. No. 9.4e-06;
Matches 280; Conservative 233; Mismatches 533; Indels 397; Gaps 60;
Qy 12 LVNPLPGPEPFTPIPTOEAAPT-----REEYEARLERYSERIWTCKSTGSSQLTHKEA 65
Db 750 MIKALEDPNLYRIGQSKIFFRGTGVAHLEERDLKI-----TDVMAFQAMCRGYLARKAF 806

Qy 66 WEEQVAVELLKEBEPFPAWYKLVLEMVHHTASLEKLVDTAMLEIMTK-----YAVGEEC 120
Db 807 TKRQOOLT-----AMKVIQRNCAAYLKLRNQWWRFLTKPKPLLIQVTRQEE 852
Qy 121 DFEVGEKMLKV-----KIVKIHPLEKVDDEATEKKS-----DGACDSPSDKENSSQ 168
Db 853 EQWAKKEEMQKITERQQAKETELKEHQHTQALAEKHTLLOLQALQAESETLYAASEEMVRV 912
Qy 169 IAQDHQKETEYVKEDEGRRESINDRARRSPRLPTSLKKGGERKWPPEKPLPHKYDVKLQN 228
Db 913 LAAKQELEELIHEARLEEBEEDRRQ-----LQAEK-----KWAQQMLDLEEQ 959
Qy 229 EDKIISNVPADSLIRTERPPNKEIVRYFIRHNALRAGTCGENAPVWVVEDELVKYKSLPSKF 288
Db 960 EE-----EEAARQKLQ-----LEKVTAAEKIKLEDDILVMDQNSKL 997
Qy 289 S---DPLDPPYKMTLNPSTKEKNTGSPDRKPSKSKTDNSLSPLNPKLWCHVHLKKS 345
Db 998 SKERKLLBEERSDLTTNLAEBEKA-----KVLTKLKSHEMSISEL-----EVLKKE 1046
Qy 346 LSG-----SPLKVKNSKNSKSPPEHLEEMKMSPNKLTNFIPIKGGPPAKPGKHSKXP 401
Db 1047 EKSROLEKLEKLESGDASDFHEQIADLQAAELKCM-----QLAKKEELQAAALRDEE 1102
Qy 402 LKAKGRS-KGI-----LNGQKSTGNSKSPK-----GLKTPKTKMOMTLTD 442
Db 1103 TAAKNALKKIRELECHTSDLOQEDLDSERAARNAKQKRODLGELEALKTELED--TLD 1160
Qy 443 MAKGTQKMTAPRNSGGT-----PRTSKPKHLPAPAAHLIAYYKENKOREDKRSAL 495
Db 1161 -STATQOELRAKREOEVTVLKALDEETRSHEAQVQEMRQKHTQAVEELTEOLEQPKRA- 1218
Qy 496 SCVISKTARLLSSSEDRARLPEELRSLVQKRYELLEHKRWASMSBEORKEYLKKKEELK 555
Db 1219 KANLDSKQTLKEN-ADLAGELVLGQAKQB-VEHKK-----KLEVOQLDQ 1265
Qy 556 KKLKAKAKERKEKEMLERLEKQRYEDQELTGKPLPAFLVDTPEGLPMTFLGDMVWVE 615
Db 1266 SKCSD--GERARAELSDKVHKLQ--EVESVTG-----MLNEAEGKAIKAKDVASLGS 1315
Qy 616 FLSCYSGILLPDAQYPIITAVSLMEALSADKGGFLVNLVILLQTLLODEAED----- 670
Db 1316 QLQDTQELLQEBETROKLVNSTKRLQLEDERNS-----LQDQLEDEMEAKQ 1360
Qy 671 -----YGBLGMKLSIPLTLHVSSELVRLCLR-RSDVQEESEGSDTDNKNDSAAFE--- 720
Db 1361 NLERHVSTINTQLSDSKKQLQDFASTIEVMEEGKRLQKEMEGLSQOYEKAAAYDKLEK 1420
Qy 721 -DNEVQDF-----LEKLTSEFFELTSEEKQLILTALCHRLMTYSVQ 763
Db 1421 TKNRLQQLDLDLVLDNQRQLVSNLEK-KQKFPDQLAAEK-----NISSKYA-- 1468
Qy 764 DHMETRQMSAEL-WKERLAV-----LKEENDKKRAEKQKREKEMAKKENGKVEN 813
Db 1469 ---DERDRAEABAREKETKALSARALFEALKELEERTNIVLKAEMEDLVSSKDDVCK 1525
Qy 814 GLGTDRKRIIVKEFPQVD-----TEAEDMISAVKSRRL-----LATOAKEREIOBR 861
Db 1526 NVHELEKSKRAL--ETQMEEMKTLQEESEDVQATEDAKLRLEVNMQALKGQFERDLQAR 1583
Qy 862 -----EMKVQLERQ-----AEEERIRKHAAAEKAFQEGYAKAKLVNRRTP-I-GTDRN 908
Db 1584 DEQNEEKRRQLRQLRQHEYTELEDERKQRAALAAAKKLEGDLKDLQADSAIKGRE-- 1641
Qy 909 HNRWLFSDVEVPGLFIEKGWVHDSIDYRFNHHCKDHTVSGDEDC-----PRSKANLGNKA 965
Db 1642 -----EATKQLRKLQAMKD-----FORELDDARASRDEIFATSKENEKKAKSLEAD 1688
Qy 966 SWNTQHGAT-----EVAVETTPPKQGNLWFLCDSQKDELINCLHPQGI 1012
Db 1689 LMLOQEDLAABARAKQADLEKEELAEELASLSRGNT--LODEKRRLE-----A 1736
Qy 1013 RESQLERLEKRYQDIHSHIHLARKNPLGLKSCDGNQELNPLRSLDIEVATRLQGGIG 1072

Db 1838 EIGAALKLEQIDELRSK-LEQEQAK-----ATQADKSKKTLEG----- 1874
Qy 1127 SAKTEVDEBEKKMVE-EAKVASALEKWKTAIREAQTFSRMHVLLGLMDACIKWMSAENA 1185
Db 1875 -----EIDNLRQIIEDEGKIWKLEKKA-----LEGELE 1905
Qy 1186 RCKVCPKGGEDDKLLILCDECNKAFHLFCIRPALYVEPDGEMOCPACQATARNRSGRNY 1245
Db 1906 ELRETVEEAEDSK-----SEAEQSKRLVEL-----ELED-----ARRNLQ----- 1940
Qy 1246 TESASEDESDDESDEEBEEDYEVAGLRL-RPRKTIRGKHSVIPPAARSGRR 1304
Db 1941 KEIDAKEIYAEADAKSNLQRIEIVAKGRLEBESARTNSDRSKRL----- 1984
Qy 1305 PGKKPHSTRRSQPKAPPVDDAEVDLVLQTKSSRSQSLQKCEBILHKIVKRYFSWPF 1364
Db 1985 -----EAEIDALTAQVDAEQKAKNOQIENKKTIELKRYKVF-- 2023
Qy 1365 REPVTRDEADYDVITHPMDFTQVQNKSCGSYRSVQBFLLTDMKQVFTNAEYVYNGRSH 1424
Db 2024 -----GESEK-----TKTKEFLVVE-KLETDYKRAKKEAADEQQQLT----- 2060
Qy 1425 VLSCMWKTQCLVLLHKLPHGYPVRRKRPKPPDLAEDGDSPEPAVGQSDRDRRR 1484
Db 2061 -----VEND-----LRKHLSEISLL-----KQADIKL-----QRDHDKTKR 2091
Qy 1485 EAEIQ 1489
Db 2092 ELETE 2096

RESULT 14
T16270
hypothetical protein F35D11.11 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 09-Jul-2004
C:Accession: T16270
R:Fulton, B.
submitted to the EMBL Data Library, June 1995
A:Description: The sequence of C. elegans cosmid F35D11.
A:Reference number: Z18487
A:Accession: T16270
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-1827 <FULL>
A:Cross-references: UNIPROT:Q20042; EMBL:U29381; NID:G868214; PID:G868224; PIDN:AAA68757
A:Experimental source: strain Bristol N2
C:Genetics:
A:Gene: CESP:F35D11.11
A:Introns: 76/2; 131/3; 159/3; 185/3; 221/3; 253/3; 320/1; 869/3; 1133/3; 1205/2; 1250/1

Query Match 3.7%; Score 295; DB 2; Length 1827;
Best Local Similarity 20.2%; Pred. No. 1.5e-05;
Matches 317; Conservative 261; Mismatches 627; Indels 366; Gaps 72;

Qy 121 DPEVGKEM-----LKVIVKIHPLEKYEDEBATEKKSQDAGCDSPSS---DKEN-SSQTA 170
Db 55 DIGVGENLDELFAKLKEBELFKNLTLEEVNEMLE-ENDAALAAHRLRVDTNLSRQLQ 113
Qy 171 QDHQKQKTYVKBEGRESINDRARRSPKLP-----TSKKGERKWAPP 215
Db 114 QLQQQOHT-----ESMRFSRNTSRELQYCKGPEICNTFLFLFSCPKYQRYRQ 164
Qy 216 KFLPHKYDVKLQNEDKIIS-----NVPADSLIR---TERPPNKEIVRFIRHNALRAGT 266
Db 165 TETQHRKLISLKEFTAVRQLHELRTTTTANDLRLQLETFRCATILMRKAIRH----- 217
Qy 267 GENAPVVEDELVKKYSPLSKESDFL---LDPKYMYTLN---PSTKKNQTSQDRK-----P 317
Db 218 -----AEQNLQDQKEQMKREKDDVLDETRLQNLNNTYMKSEKANERQDLKKEDE 271
Qy 318 SKSKSTDNSSLSPLN--PKLWCHVHLKKSLSGSLPKVKNKSNKS-----PEEHLEMM 370

Db 272 CRKLEQNDDELSDILEQLSKMAHEMAGGRGRSQSPSVIVDNTTAHSLFNPETPMDVARKWR 331
Qy 371 KMSPNKLTHTNPHIPKGPAPKPKGHSKPL-----KAGRSKGILNGQKSTGNSPKK 426
Db 332 KULTTK-----NEIDESRAAKQAERDRACKDLEKEBKRRKDDREARSKRSVTVSRE 387
Qy 427 -GLK-----TPTKMKQMTLLDMAGTQQTQVTRAPRNSGTPRTSSKPHKHLPPAALHLTA 480
Db 388 HDLKKLDDLELRKASEKIRNLBEQRESQEKLTISVQNS---LNEAHRQHQFIEEL--MIR 442
Qy 481 YYKKNKDRDKRSALSCVISTARLLSSDRARLPEEL--RSLVQKRYLLEHKKR----- 534
Db 443 HREELKEREDSH-----EEALRSKQDTEERSFEKERSERKIRRESDELRETQSLKG 495
Qy 535 -WASSEE-----QRKEYLKKREELKKLKEKAK--ERREKEMLE-RLEKOKRYDQDEL 585
Db 496 DVAAKMTDLDKTLKLDMLTEDEBELKKLETEREQADQORDLEIAECRAKLEDEMAKEAE 555
Qy 586 TGNLPAFLVDTP---EGLPNTLFGDVAMVVFSLSCYSGLLLPDQAPYITAVSLMEALS 642
Db 556 LRKELAEFOAILITAMEGEGKLNQ-----EQFLESKNEL-----NTLTDQIESLN 599
Qy 643 ADKGFLYINRVILVILLQLODEI-----ADYGEELGMKLSIEIPITLH 686
Db 600 SEVEN---KNEIRINLMATLQEKVHIQNVRTSSHOLTATYBEANGEIDILKAELTRLHE 656
Qy 687 SVSELVRLCLRESDVQESSESDTDNDKDSAAFEONEVQDFLE--KLTSFPELTSEE 744
Db 657 QVNE-----RTRQISEANEKYDDAARKNDALLEVATWQEKYQQLKMLEENRRGQEK 710
Qy 745 KIQ--ILTALCHRIILMTYSVQDHMETROQMSAELMWERLAVLKEENDKKRAEKQKKEM- 801
Db 711 EREEADLRALLDRLGNFDPKLTNELKQKGVTVDSLNEEISSLKEQLNK--SEKERKEELL 768
Qy 802 -----EAKNGKGVENGIGTKTKRIRIVKTFEPQVDTBAEDMISAVKSRRLA--- 849
Db 769 RMEELQKNEABEMKEEYEVKLQLAEKDR-QGVENFGCECEARMNEL---TKIHEMLMEEH 824
Qy 850 -----IQAKKERIQREREMKVLER-----QABEERIRKHKAAAKAFQSG 890
Db 825 DQKVDHLHTEEEVERLKEKMRKELEKLEQNEQNGDRAEWSNRNRNLESSKNEAVTELOER 884
Qy 891 TAKAKLVMRRTPIGTDRNHRNYLFSDEVPGLFIEKGWVHDSIDYRFNHHCDDHTVSGD- 949
Db 885 VQKLEDDVVK-----EKEDKETALARD-----LEDS-----HEKSRDL 916
Qy 950 EDCPRSKKANIGKNASMNTOHGTATEVAVETTPKQGNLWFLCDSQKELDELNLCLHP 1009
Db 917 DDKLRKMLTDBEKEEDRCKEOKTLNEERMKLM-----EQKEEAMLVATKHA 963
Qy 1010 QGIRSQLKERLEKYQDIIHSIHILARKPNLGLKSCDGNQELLNLRSLIEVATVLOKQ 1069
Db 964 TTI--DQOTRISVLQGV-----EKLTAGIAE---RESSINALESTMLISKL---- 1008
Qy 1070 GLGYVEETSEFEARVISLEKLADFGCEVIAL-QASVIKKFLOGFMAP-KQKERKLOQSEDS 1127
Db 1009 -----ETEA-----LEKLD--ELAVMLQNSLKNGLKGLSEKWNNEERKKIQD--- 1052
Qy 1128 AKTEVDDEKKNVBEAKVASA-LEKWKTAIREAQT--FSRMHVL-LGMLDACIKWMSAE 1183
Db 1053 -LADQLREANKVHVNMKNVLEKKNELDQNVDTLTKVRQLEIQLMDKAAKNEVSGD 1111
Qy 1184 NARCKVCPKGGEDDKLLILCDECNKAFHLFCIRPALYVEPDGEMOCPACQATARNRSGR 1243
Db 1112 LLR-----KMEHDAOSMLKQAQNEQFRJTDLEKVRKALQD--ENQRLVNDLTVKAAFEVK 1165
Qy 1244 NYTESASGSDSDDE 1291
Db 1166 RETSKAISLDIKYRSAEKANKGELDNQRLSRDLATVTLKLE-RQELKAKDSNRLRD 1224
Qy 1292 -----HSTVPPAARSGRRFGKPKHSTRRSQPKAPPV-----DAEVDLVLQ 1333

Db 1225 SOKRFEVQSKLANLQKSAVESLONPMSSNSRQNRSIYVDIPRAASSICLNENSDEVLRL 1284
QY 1334 TKRSR--RQSLQKCEILLHKI---VKYRFSPPFPVTRDEAB--DYVDVITHPMD 1385
Db 1285 SPSVRFADSSQNMQRADVMDVSSVGVTLRF---LKERIBQLEADNADISDALEKAKD 1341
QY 1386 FOTVQNKSCG-----SYRSVQBFLLDMKQVF--TNAEVNCRGSHVLS 1427
Db 1342 ELQRNEKLADROMVIERVEROLVHI TEERNTIENRMTSQRQMYLTNESSRSRREHRS 1401
QY 1428 CMVTEQCLVLLHKLPHGVYRKRKFPDRAB-----DEGDSPEAVGQSRD 1478
Db 1402 MKAR-----ISTLELHL-----REKESKLAHLRKEIEVLVHQLHDALESKEKATGLGV 1450
QY 1479 EDRRSREAEIQ 1489
Db 1451 QDSKRDLEEQ 1461

RESULT 15
JC5837
364K Golgi complex-associated protein - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 05-Mar-1998 #sequence_revision 13-Mar-1998 #text_change 09-Jul-2004
C:Accession: JC5837
R.Toki, C. Fujiwara, T. Sohma, M. Hong, H.S.; Misumi, Y.; Ikehara, Y.
Cell Struct. Funct. 22, 565-577, 1997
A:Title: Identification and characterization of rat 364-kDa Golgi-associated protein re
C:Superfamily: giantin
F:49-549, 624-1176, 1238-1707, 1763-3114/Domain: coiled-coil leucine zipper #status predic
F:3165-3187/Domain: membrane anchor #status predicted <MAD>

Query Match 3.6%; Score 290.5; DB 2; Length 3187;
Beat Local Similarity 18.3%; Pred. No. 4.6e-05;
Matches 350; Conservative 266; Mismatches 676; Indels 625; Gaps 67;

QY 19 REPFTIHTQAFRTREYERARLERYSERIWTCKS--TGSQSLTHKEAWEEQEVAELL 76
Db 919 EEPVCKEALQOELEWLKRESEQRKQLQAALISRKELLQKSVKLEELAKVREESTKDSL 978
QY 77 KEEFPAPWYKLVLEMHVHTASLEKLVDTAW--LEIMTKYAVGE--ECDPEVGKEKMLVKV 133
Db 979 RES-----EKRELEDSKNKODPEKYGTSEWRELEVSRLTISEKEVELE-GIRDLAK 1032
QY 134 IV-----KTHPLEKVDDEATEKKSD-----GACDSPSSDKENS 166
Db 1033 AAABEELQALVQMTQDLQNTKQIDLLQEEITENQATIQPITGTMDAGDGSVAKETS 1092
QY 167 SQI-----AQDHQKETEYVKEDEGR-----RESINDRARRSPRK 200
Db 1093 VSSPPRAGGGEHWKPEL-----EGKIVDLKEKTLQKLOKQALISRKAILKKAQEKKH 1147
QY 201 LPTSLKGERKWAPPKFLPHKYDVVLQNEDKLISNVPADSLIRTERPKNKEIVRYFIRHN 260
Db 1148 LKEELKEQDAY--RHLEQFDGGSKENI-----RAPRLQAKESTDQ 1191
QY 261 ALRAGTGNAPWVBEVLVKYSLSPKFSDF-----LLDPYKY 298
Db 1192 QL-PGTGQOEPHSGSEGLSEGTPEASESDLHAAQSPHGETATLQATVVAQIQDLKE 1250
QY 299 MTLNFTKKNKGTSPDRPESKS-----KTDNSSLSPLNPKLWCH-----VHLKKS 345
Db 1251 IEVEKEELELKTSSTTSLTCKSEVLLQBOINEQGLEIQNLKAASHEAKAHTQLKOE 1310
QY 346 LSGSPKLVKNSKNSKSPSEHLEEMKMSPNKLHNFHIPKKGPPAK----- 393

Db 1311 LESSQLIADLEHLKTLQPELLETLQKHVGQKEEVSYLVGQGEKEQTLTTVTQEMEQE 1370
QY 394 ---PKHSDKPLKAKGRSKGILNGQKSGN--SKSPKGLKTPKTKMKTMLDMAKGTQK 449
Db 1371 RLIALHTQLEMQAKEHEERLKQVQVEICELKQPKLEEBESKAKQQLQRLQAAALISRK 1430
QY 450 MTRAPRNSGGTPTSSKPH--KHL-----PPAALHLIAYYKENKDR--- 488
Db 1431 EALKENKSLQQLSSARDAVEHLTKSLADVESQVSQVQNEKDALGLKALQEEERDKLIV 1490
QY 489 -----BDKRSALSCVISKTARLLSSBDRARLPEELSL-----VOK 524
Db 1491 EMDKSLLENQILGGSCESLKLALGGLTDEKELKKELESVRCSKIAESTEWQEKHELOK 1550
QY 525 RYELL-----EHKRW-----SMSEQORKEYLKKREELKKLKEKAKERRREKMLE 572
Db 1551 EYEVLLQSYENVNEAERIQHVESVRQOEYVAKLSAESDKREERKQLQDAEQEMEE 1610
QY 573 RLEKOKRYEDQ-----ELTGKNLPAFLVDTPEGLPN-----TLFGDVAMVVPEFLSCYS 622
Db 1611 MKERKRFPAKSKQOKILEEEN--DRLRAEAQPVGGANESMEALLSSNASLKEELERIT- 1668
QY 623 LLLPDAQYPTTAVSLMEALSADKGGFLYLNRLVLI-----LLQTL----- 663
Db 1669 ---LEVK-TLSKEFEALMAEKNTLSBETRNKLQVEAQELKQASLETTEKSDPKDVI 1722
QY 664 -----QDETAEDYGEIGMKLSEIPLTI-----HSVSLEY----- 692
Db 1723 EEVTEAVVGKSQEQDSLEEN-----AKLEDAEATLLANSAPKPGVSETFSSHDDINNYYLQ 1777
QY 693 ---RLCLRRSDVQESSEGSTDDNDKSDAAFNEDNEVQDEFLEKLE----- 733
Db 1778 LDQKGRITAELEMSKQKDRLESQTLNEKKNALLTQISAKDSELKLESEVAKINMLNQI 1837
QY 734 -----TSEFFPLESEELQILALTCHRI-----LMTYSVQDHPMTR 769
Db 1838 QEELSRVTKLKTAEAEKDDLEERLMNLQBELNGSIGNYQDVDTDAQIKNEQLESEMQL 1897
QY 770 QOMAEALWKEKRLAVLKEENDKKRAEKQKKE-----MEAKN 805
Db 1898 KRCVSELEERQQLVKE---KTKVESEIRKEYMEKIQGAQKPGSKIHAKELOELLKEKQ 1954
QY 806 KENGKVENG---LGKTDKRRKRVKFPQVTEAEDMISAVK----- 843
Db 1955 QEVQQLQKDCRYLQRLQISALEKTVALEFVITESQKOLDATKGNLAQAVEHHKKAQELS 2014
QY 844 -----SRRLLAIQAKEREIOEREMVKLE--RQAEERIRKHKAAAEK----- 885
Db 2015 SFKILLDDTQSEAAARVLADNLKKEKQSNKESIKSQIKQKDEDLRLLEQAEKHKKEK 2074
QY 886 -----APQEGTAKAKLVNRRTPIGTDRNHNRYWLFSDVPGILF 923
Db 2075 KNMOEKLDALHREKAHVEDTLAEIQVSLTRKDKOMKELQOQLSDSTLAQAAFTKSMSSLQ 2134
QY 924 1EKGWVHDSI--DYRFNH--HCKDHTVSGDEDEYCPRSKKNALGNKNSMNTQHGATVA 978
Db 2135 DDDRVIDEAKKWEQFGDAIQTKKEEVRLEENCTALK-----DQLQMTIH--MEELK 2187
QY 979 VETTPKQGNLW-----FLCDSQKELDELL 1004
Db 2188 ITVSRLEHDKIWEKSAQTELQHQOKYADVKLOEENKELMSQLEBAGQLYHDSKNELTKLE 2247
QY 1005 NCLHPQGIRSQLERLEKRYODIHSIHLARKNPLGLKSCDGNQELNLFNRSDLI---E 1061
Db 2248 SELSKLQDQSDTLKNSLEKREHNNLEGIITKQOEADIQNCKFNCEQ---LETDTATASRE 2304
QY 1062 VATRLQKGLGVSETSEFEARVISLEKLKDFGECVIALQASVTKKFTQGFMAPKQKRRK 1121
Db 2305 LTTLLH-----DEINVKEQKIIISLSGKE-----EAIQVAIAELHQHQSKEIKLENL 2352
QY 1122 LQSDSAAKTEVDEBEKMWEEA-KVASALEKWKTAIREAQTFSRMHVLLGLMDACIKWDM 1180
Db 2353 LSQBEENLTLEENKRAVEKTNLTLEATI KKESELEQ-----AQLDSEVFKSMS 2403

QY 1181 SAENARCKVCPKGGEDDKLILCDENKAFHFLCLRPALYVDPGEWQCPACOPATARRNS 1240
DB 2404 SLQDDR-----DRIV-----S 2414
QY 1241 RGNNTYTESASEDESEDEEBEEDYEVAGLRPRKTIRGHKSHVIPAAR 1300
DB 2415 DYQLEERHLSVILEKDELIDQAAENKKE-----BIRGR----- 2452
QY 1301 SRRPGCKPHSTRSQKAPPVDDAEVDEL-----VLQTKSSRRQSLE--LQKCEIL 1352
DB 2453 -GHMDDLNGENAKL-----DABLIQVRRDLNEVITIKDSQQRQLLEAQLQONKELR 2502
QY 1353 HKIVKRFSPREPVRDEADYDVIITHPMDFTQVQNKSCGVSVRVQEFITDMKQVF 1412
DB 2503 NECVK-----LEGRLKSEAK--OSLQMSLDALQENQGLSKETKSPKEQILTALH--- 2551
QY 1413 TNAEVNCRGSHVLSWVTEOCLVLLHKLPGHPYVRRKRKFPDRLAEDGDSPEA 1472
DB 2552 -----EGALAVYHAQL-----RVREBEVQKLTA 2576
QY 1473 VQCSRDEDRSRAEIQEWL-----QDTSYSAKINSK-----DHNCFMMLVNTQ 1517
DB 2577 LSSS-----QKRTVDLQEEVLCVQKEASKVSEIEDKLKRELKHLHNAIGIRNETE 2628
RESULT 16
T13564
microtubule-associated protein homolog - fruit fly (Drosophila melanogaster)
N:Alternate names: hypothetical protein EG:49B4.1
C:Species: Drosophila melanogaster
C>Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 09-Jul-2004
C:Accession: T13564
R:Spanos, L.; Papagiannakis, G.; Siden-Kiamos, I.; Louis, C.
A:Submitted to the EMBL Data Library, April 1999
A:Description: Sequencing the distal X chromosome of Drosophila melanogaster.
A:Reference number: Z17689
A:Accession: T13564
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-5327 <SPA>
A:Cross-references: UNIPROT:076891; EMBL:AL031128; PIDN:CAA20006.1
C:Genetics:
A:Cross-references: FlyBase:FBgn0025392
A:Introns: 24/2; 52/3; 104/3; 179/1; 232/1; 1669/3; 2566/1; 4798/3; 5272/1
A>Note: EG:49E4.1
C:Superfamily: Drosophila 576K microtubule-associated protein homolog
Query Match 3.6%; Score 290.5; DB 2; Length 5327;
Best Local Similarity 20.2%; Pred. No. 8.3e-05;
Matches 274; Conservative 213; Mismatches 491; Indels 381; Gaps 62;
QY 115 AVGECDFPVGKMKLVKIVKHPLEKVDREATEKSGDCDSSPSDKENSSQIAOHO 174
DB 2695 SVVESVKDEHDAERRESIAKV---ESVIDE-----GKSDSKS-----SQDSQ 2727
QY 175 K--KETVVKEDGRRESI-----NDARRSPRLKPTSLKGGKRWAPPKFLPHKYDVKLQN 228
DB 2728 KDEKSTLASKEASRESVSVESKDDAKESRPESVIASGEP-----VPRESKPLDS 2780
QY 229 ED-----KLIISNPADSLIRTPRNKEIVYFIRNALRAGTGNA--PWVEDELVK- 280
DB 2781 KQTSRPGSVVSVTAED-EKSEQQRRRESVAESVADTKKGQSQEASRPSSV-DELLKD 2838
QY 281 -----KYSI-----PSKFSDFLDLPYKMTLNPSTKTKNT-----G 311
DB 2839 DDEKQESRRQSTGSHKAMSTWGDSPMDKADSKSPRPSVAESI KHENTKDEESPLG 2898
QY 312 SPDPRKPSKKKTD--NSSLSPPNPKLWCHVHLKSLGSLSPKLVXNKNKSPPEEHLNEM 370
DB 2899 SRRDSVAESIKSDITKGEKSLPSK---EVSRPESVGS---IKDEK-AESRRESVAESV 2951
QY 371 KMSPNKLHTNTHIPKKGPPAKKPGKHSKDKPLKAKGRSGKILNGKSTGNKSPKGLKT 430

RESULT 17

DB 2952 KPES-----SKDATSAPPSKEHRSPEVLSGLKD--EGDKTTTGRRVSVADSIKD 2998
QY 431 PKT-----KXQMTLLDMAGTKQMTAPR-----NSGGTPTRTSKPHKLPPAA 475
DB 2999 EKSLLVQSASPESEAESELKDAAPSQTSRPESVTSVKDGKSPVASKEASR---PAS 3055
QY 476 LHLIAYYKKNKREDKRSALSCVISTARL-----LSSEDRARLPEE--LRSLVKRYE 527
DB 3056 VAENA--KDSADESKQRPESLPQSKAGSIKDEKSPASKDEAKESKESRRESVAEQFP 3113
QY 528 LLEHK-KRWASMSBQRKEYLKKRE-----ELKKLKEKAKERRKEMLERL 574
DB 3114 LVSKEVSRPASVAESVKDEAKESKESPLMSKEASRPASVAGSVKDEAKESKESRRESV 3173
QY 575 EKQKRYEDQLTGKLPAPRLVDTPEGLPNTLFGDVAMVVEFLSCYSG-----LLLPDAQ 629
DB 3174 AEKSPSPSK-----ASRPASVAESVKDEADKESRRESGAEKSPASKEAS 3222
QY 630 YPIT-AVSLMEALSADKGGFLYLNRLVILLQDEIAEDYAGELGMKLSIPLTLHSV 688
DB 3223 RPASVAESIKDEAKESK-----ESRRESVAE-----KSPSPSKEA 3258
QY 689 SELVRLCLRRSDVQSESESGDTHDKNDKSAFED-----NEVQDFLEKLET 734
DB 3259 SRPTSV---KSVKDEAKESKESRSDSVAKSPASKEASRPASVAESVQDE-AEKSKE 3314
QY 735 SEFFELTSEKLIQILTALCHRIILMTYSVQDHME--TROQMSAELWKEARLAVLKEENDKRA 793
DB 3315 ESRRESVAESKPLAYKEASRPASVAESI KDEAKESKESRRESVAESKSPASKEASRPST 3374
QY 794 EKQKREKMEAKNKENGKVENGLGTRDKRIIVKFFQVDTEAEDMISAVKSRLLAIOAK 853
DB 3375 VAESVKDEAKESKES-----RDSVAESKSPASKEASRPASVAES--VQDEAE 3421
QY 854 KERETQEREMKVKLRQAEERIRKHAA-----AEKAFQGIKAKAKLVMRTPTGTDR 907
DB 3422 KSKEESRRESVAESKSPASKEASRPASVAESVKDDAKESKEE--SRRESVAESKSPAS- 3478
QY 908 NHRNWLFSDEVPGLFIEKGWVHDSIDYRFNHCHDKDHTVSGDEDYCPRS--KXANL-GKN 964
DB 3479 EASRPASVAESV-----KDEAKESKESRRESVAESKSPSPSK 3516
QY 965 ASMTQHGHTATVAVAVETTPKQGNLWFLCDSQKELDELLN--CLHPQGIRESQLKERLE 1022
DB 3517 ASRPT--SVAESVKDEAKESKESRR---ESVAESKSLASKEASRPASVAES-VKDEAE 3569
QY 1023 KYQDIIHSIHLARPNLGLKSCDGNQELLNPLRSDLIEVATRLQKGGLYYEETSEFEA 1082
DB 3570 KSKEE-----SRRESVAESKSPASKE-----ASRPAS-----VAESVKDEA 3605
QY 1083 RVISLEKLKDFGECVIALQASVVIKKFLOGFMAPKQKRKLQSEDSAKTEEVDEEKKWVE 1142
DB 3606 -----EKSE-----VSRRESVAEK--SPLPSKEASRPTSVAESVK-----DEADKSKBE 3648
QY 1143 AKVASALEKWKTAIREAQTFSRMHVLLGLMDLADCIKWDMSAENARCKVCPKGGEDDKLILC 1202
DB 3649 SRRESGAEKSPASMEASRPTSV-----AESVK----- 3676
QY 1203 DECNKAFHLFCLRPALYVDPGEWQCPACOPATARRNSGRNYTEES-----ASDSED 1256
DB 3677 DETEX-----SKEESRRESVTEKSPSPSKESRPTSV 3708
QY 1257 DESDEEBEEDYEVAGLRPRKTIRGHKSHVIPAARSGRRP----- 1305
DB 3709 AESVKDEAKESKESR-----RESVAESK-----PLASKESRSPASVAESIKDEA 3753
QY 1306 -GKKPHSTRSQKAPPVDDAEVDELVLQTKSSRRQSL 1343
DB 3754 EGTQESRRSNPESGKAESIKGQSSLSAKETSRRPDSV 3792

JC5421
smooth muscle myosin heavy chain 2 - mouse
C:Species: Mus musculus (house mouse)
C>Date: 10-Jun-1997 #sequence_revision 18-Jul-1997 #text_change 09-Jul-2004
C:Accession: JC5421
R:Hasegawa, K.; Arakawa, E.; Oda, S.; Matsuda, Y.
Biochem. Biophys. Res. Commun. 232, 313-316, 1997
A>Title: Molecular cloning and expression of murine smooth muscle myosin heavy chains.
A:Reference number: JC5420; MUID:97242182; PMID:9125171
A:Accession: JC5421
A:Molecule type: mRNA
A:Residues: 1-1938 <HAS>
A:Cross-references: UNIPROT:O08638; DDBJ:D85924; NID:G1945079; PIDN:BAAL19691.1; PID:G194
A:Experimental source: smooth muscle
C:Comment: This protein plays a role in smooth muscle cell contraction.
C:Superfamily: myosin heavy chain; myosin motor domain homology
C:Keywords: nucleotide binding; P-loop
F:88-771/Domain: myosin motor domain homology <MMOT>
F:178-185/Region: nucleotide-binding motif A (P-loop)

Query Match 3.6%; Score 289.5; DB 2; Length 1938;
Best Local Similarity 20.0%; Pred. No. 2.8e-05;
Matches 259; Conservative 224; Mismatches 499; Indels 315; Gaps 57;

Qy 12 LVNPLPGGEPPPTIPIHTQEAFT-----RREYEARLERYSERIWTCKSTGSSQLTHKEA 65
Db 750 MIKALELPNLYRIGQSKIFRTGVLAHLEERDUKI---TDVWAFQAMCRGYLARKAF 806
Qy 66 WEEEOEVAELKEEFPWYKLVLMVHHNTASLEKLVDTAWLEIMTK-----YAVGEEC 120
Db 807 TKRQOQLT-----AMKVIQRNCAAYLKLNNQWWRFLTQVKPLQVTRQBE 852
Qy 121 DFEVKEKMKLV-----KIVKIHPEKVEDEATEKKS-----DGACDSPSKENSSQ 168
Db 853 EQMAKEEEMQKITERQQAETELKEQHTQLAEEKTLQBLQEAETELVAESEMVR 912
Qy 169 IAQDHQKETVYKDEGRRESINDRARSPRKLPTSLKKGKRWAPPFLPHKYDVQLQN 228
Db 913 LAACKQELIEIHEARLEEEEDRRQ-----LQAEKK-----KMAQOMLDESQL 959
Qy 229 EDKIISNVPADSLIRTPPNKEIVRYFIRHNALRAGTGENAPWYVEDELVKYSLSPKF 288
Db 960 EE-----EEAARQKLQ-----LEKVTAEAKIKKLEDDILVMDQNSKL 997
Qy 289 S----DFLDPYKYMTLNSTRKKTGSDPRKPSKSKTDNSLSPLAPKWLCHVHLKKS 345
Db 998 SKERKLLREVSDDLTLNLAEEBEKA-----KNLTCLKSKHESMISEL-----EYRLKKE 1046
Qy 346 LSG-----SPLKVNKSNKSPPEHLEEMMMKMSPNKLHTNFHI PKKPPAKPKGKSDKP 401
Db 1047 EKSROLEKLKRLKLEGDASDFHEQIADLQAIKLM-----QLAKKEELQALARLDEE 1102
Qy 402 LKAKGRS-KGI-----LNGQKSTGNSKSPK-----GLKTPKTKMKQMTLLD 442
Db 1103 IAKKNALKKIRELEGHISDLQEDLDSERAAKNKAEKQKRDIGEELEALKTELED--TLD 1160
Qy 443 NAKGTQKMTAPRNSGGT-----PTSSKPKHLPPAALHLIAYYKENDREKRSAL 495
Db 1161 -STATQELRAKREQEVTVLKALDEETRSHEAQVQEMRQKHTQAVEELTEQLEQFKGA- 1218
Qy 496 SCVISKTRALLSSEDRARLPBELRSLVOKRYELLEHKRWASMSSEQRKEYLKKREELK 555
Db 1219 KANLDSKQTLKEN-ADLAGELRVLGAKQB-VEHKKK-----KLEVLQDLQ 1265
Qy 556 KKLKEKAKERREKEMLERLEKQRYEDQELTKNLPFAFLVDTPPEGLPNTLFGDVMVVE 615
Db 1266 SKCSD--GERARALSDKVHKLQN-EVESVTG-----MLNEASGKAIKLAKDVASLGS 1315
Qy 616 FLSCVSGILLPDAQYPTAVSLMELASDKGFLYLNRLVILLQTLQDETAED----- 670
Db 1316 QLQDTQELQEBETROKLANVSTKLRLDEERN-----LQDQLDEEMAKQ 1360
Qy 671 ----YGELGMKLSIPLTLHVSSELVRLCLR-RSDVQESSEGSDDDNKDSAAFE---- 720

Db 1361 NLERHVSITNLQSDSKKKLQDFASTIEVMEEGKKRLQKEMEGLSQQVEEKAAAYDKLEK 1420
Qy 721 -DNEVQDFE-----LEKLTSEFPFELTSEBKLLQILTALCHRLMTYSVQ 763
Db 1421 TKNRLQELDDLVDLQNRQVSNLEK-KOKKFDQLLAEEK-----NISSKYA-- 1468
Qy 764 DHMETRQMSAEL-WKERLAV-----LKEENDKKRAEKQKKEKMEAKKENGKVEN 813
Db 1469 ---DERDRAEAAREKETKALSARALEEALEAELEERTNKLMAEEDLVSSKDDVGK 1525
Qy 814 GLGKTDKRRKRVKPEQVD-----TEADMISAVKSRL-----LAIQAKKEREIOER 861
Db 1526 NVHELEKSKRAL--ETQMEEMKTQLEESDDVQATEDAKLRLEVMNQALKGQFERDLQAR 1583
Qy 862 -----EMVKULERO-----AEBERIRKHAARAKAFQEGIAKAKLVMRTPPI--GTDNR 908
Db 1584 DEQNEEKRRQLQRQLHEYTELEDERKQRLAALAAAKKLEGLDLKDLQADSAIKGRE-- 1641
Qy 909 HNRWLSFDSVPGGLFIEKGWVHDSIDYRFNHHCXDHVTYSGDEDYC---PRSKKANLGKNA 965
Db 1642 -----EAIQURKQLQAQMKD-----FORELDARASRDEIFATSKENEKAKSLEAD 1688
Qy 966 SMNTQHGTTAT-----EVAVETTTTPKQGNLWFLCDSQKDELNCLHLPQGI 1012
Db 1689 LMQLOEDLAAARARKQADLEKEELAEELASSLSGRNT--LQDEKRRLE-----A 1736
Qy 1013 RESQLERLEKRYODIHSIHARKPNGLKSCDGNQBELNPLRSLDLIEVATRLQKGGIG 1072
Db 1737 RIAQLEEELEBEQGNMEAMSDRVKATLQA-----EQLSNELATER-STAQKESARQQ 1789
Qy 1073 YVEETSFEARVISLE-----KLKDFGCVIALQASVTKKFLQGFMAPKQKE--RKLQS 1124
Db 1790 LERONKELRSLKQVEGAVKAKLK---STVALEAKIAQLEBEQVEQEAKEQAATKSLKQ 1846
Qy 1125 EDSAKTE---EVDBEKQNVBEAKVASALEKWKTAIRE 1158
Db 1847 KDKKLKEVLLQVEDERKVAEQYK-EQAEKNGTKVKQ 1881

RESULT 18
A45592
C:Species: Plasmodium falciparum
C:Date: 22-Nov-1993 #sequence_revision 02-Dec-1994 #text_change 09-Jul-2004
C:Accession: S24597; A45592; S29393; S34842; B45592; C45592; D45592
R:Zhu, J.; Hollingdale, M.
submitted to the EMBL Data Library, November 1990
A:Reference number: S24597
A:Accession: S24597
A:Molecule type: DNA
A:Residues: 1-1909 <ZHU>
A:Cross-references: UNIPROT:Q25893; EMBL:X56203; NID:G9915; PID:G9916
R:Zhu, J.; Hollingdale, M.R.
Mol. Biochem. Parasitol. 48, 223-226, 1991
A>Title: Structure of Plasmodium falciparum liver stage antigen-1.
A:Reference number: A45592; MUID:92107224; PMID:1840628
A:Accession: A45592
A:Molecule type: DNA
A:Residues: 1-195;638-688;1165-1215;1590-1909 <ZHU>
A>Note: sequence extracted from NCBI backbone (NCBIN:75010, NCBIN:75012, NCBIN:75014, NC
R:Guerin-Marchand, C.; Druilhe, P.; Galey, B.; Londono, A.; Pataraotikul, J.; Beaudoin,
Nature 329, 164-167, 1987
A>Title: A liver-stage-specific antigen of Plasmodium falciparum characterized by gene
A:Reference number: S29393; MUID:87315391; PMID:3306406
A:Accession: S29393
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 323-387 <GUE1>
A:Cross-references: EMBL:M28266
R:Guerin-Marchand, C.; Druilhe, P.; Galey, B.; Londono, A.; Pataraotikul, J.; Beaudoin,
submitted to the EMBL Data Library, April 1992
A:Description: a liver-stage-sepcific antigen of plasmodium falciparum characterized by gene

A:Reference number: S34842
 A:Accession: S34842
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 323-381, 'HKA1', <GUE2>
 A:Cross-references: EMBL:W28266
 A>Note: difference at carboxyl end due to frameshift error
 C:Comment: This protein is found as flocculent material in the parasitophorous vacuole.
 C:Superfamily: trichoyalin; calmodulin repeat homology
 C:Keywords: EF hand
 F:154-1629/Region: 17-residue repeats (A-K-E-K-L-Q-E-Q-Q-S-D-L-E-Q-E-R-R)

Query Match 3.6%; Score 287.5; DB 2; Length 1909;
 Best Local Similarity 18.5%; Pred. No. 3.4e-05;
 Matches 296; Conservative 282; Mismatches 545; Indels 477; Gaps 65;

```

QY 36 EYEALE---RYSERIMTKSGSSQLTHKAWEEQEAELKEEPPAWKEKLVLEMY 92
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 295 QEQQSDLEERRAKEKLEQQSDLEQERLAKEKLEQQSDLE-----QERLAKEKL 345

QY 93 HNTASLEKLVDTAMLEIMTKYAV-GECDFFVGKMKLVKIVKIH-PLEK---VDEEA 147
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 346 QEQQSDLEQ-----ERLAKEKLEQQSDLE--QERLAKEKLEQQSDLEQRLAKEKL 396

QY 148 TEKSDGACDSFSSDK--ENSSQIAQDHQK-----ETVVKEDEGRRSINDR 193
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 397 QEQQSDLEQERLAKEKLEQQSDLEQERRAKEKLEQQSDLEQERLAKEKLEQQSDLEQ 456

QY 194 ARSPKPLTSLKGERKWA PKPLPHKYDVKLQNEKDIISNPADSLIRTE-RPNKEI 252
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 457 ERRAKEKLEQQSDLEQ-----RRAKEKLEQQ-----SDLEQERLAKEKLEQQSDLE 505

QY 253 VYFIRHNAIRAGTGENAPWVEDELVKYISLPSPKFSDFLLDPYKMTLNPSTKKNVTS 312
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 506 EQERLAKEKLEQQSDS-----EQERLAKEKLEQQSDL-----E 540

QY 313 PRKPSKSKTNSLSPLNPKLMCHVHLKXLSGSPKLVKNS--KNSKSPHEHLEMM 370
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 541 QERLAKEKLEQQSDLE-----QERLAKEKLEQQSDLEQERLAKEKLEQQ 587

QY 371 KMSPNKLTNTHI PKGPPAKPGKHSDKPLKAGRSKGIINGQKS-TGNSKSPKGLK 429
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 588 SLEQERL-----AKEKLEQQSD--LEQERLAKEKLEQQSDLEQERLAKEKLEQ 635

QY 430 TPKTKMOMTL-----LDMAKGTQKMTAPRNSGGTPRTSSKPKHCLPPAALHLIAYKEN 485
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 636 EQQSDLEQERLAKEKLEQQSDLERTKASK-----ETLQEQSDLEQERLAKEKLEQQ 689

QY 486 KREDKRSALSCVISKTA RLSSEDRA--RLPEELSLVQKRYELLEHKRWASMSERQ 543
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 690 SLEQERRAKE-KLEQQSDLEQERRAKEKLEQQSDLEQERRAKEKLEQQSDLEQERR 748

QY 544 -KEYLKKKEEL-----KKLKEK-----AKERREKML-----ERLEKQRYEDQ 583
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 749 AKEKLEQQSDLEQERLAKEKLEQQSDLEQERRAKEKLEQQSDLEQERLAKEKLEQQ 808

QY 584 ELTKGNLPAFLVDPTEGLPNTLFGDVAMVVEFLSCYSGLLPDQYPIYAVSLMEALSA 643
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 809 S-----DLEQERRAKEKLEQQS 826

QY 644 DKGFLYLNRVILLQTLLODEIADYDYGELGMKLSIEPLTLHVSSELVRLCLRSVDQ 703
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 827 D-----LEQERLAKE-----KLQE-----QQSDLEQ 847

QY 704 ESEGS-----TDNKSAAEFEDNEVQDFLEKLETSEFFELTSEKLIITLALCHRL 757
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 848 ERRAKEKLEQQSDLEQERLAKEKLEQQSDLEQ-----ERRAKEKLEQ----- 890

QY 758 MTYSVDHMETR-----QMSAELWKEKRLAVLKEENDKKRAEKQKR----- 799
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 891 ----EQQSDLEQERRAKEKLEQQSDLEQERLAKEKLEQQSDLEQERRAKEKLEQQSD 947

QY 800 EMEAKNKENGKVENGLGKTRKRIKVPFQVDTEA-----DMISAVKSRLL--A 849

```

RESULT 19

S03166
 myosin heavy chain, gizzard smooth muscle [similarity] - chicken
 N:Contains: myosin ATPase (EC 3.6.4.1)
 C:Species: Gallus gallus (chicken)
 C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 09-Jul-2004
 C:Accession: S03166; A27066; A26045; A36604; A43298
 R:Yanagisawa, M.; Hamada, Y.; Katsuragawa, Y.; Inamura, M.; Mikawa, T.; Masaki, T.
 J. Mol. Biol. 198, 143-157, 1987
 A:Title: Complete primary structure of vertebrate smooth muscle myosin heavy chain deduced from complementary DNA
 A:Reference number: S03166; MUID:88118918; PMID:2892941
 A:Accession: S03166
 A:Molecule type: mRNA
 A:Residues: 1-1979 <YAN>
 A:Cross-references: UNIPROT:P10587; EMBL:X06546; NID:G63633; PIDN:CAA29793.1; PID:G63633
 A>Note: part of this sequence was confirmed by protein sequencing
 R:Maizumi, T.; Onishi, H.; Yajima, E.; Matsuda, G.
 J. Biochem. 102, 133-145, 1987
 A:Title: Amino acid sequence of the amino-terminal 24 kDa fragment of the heavy chain of chicken gizzard smooth muscle myosin
 A:Reference number: A27066; MUID:88032919; PMID:331284
 A:Accession: A27066

A:Molecule type: protein
 A:Residues: 2, 'Z', 4-204 <MAI>
 J:Onishi, H.; Maita, T.; Miyaniishi, T.; Watanabe, S.; Matsuda, G.
 J. Biochem. 100, 1433-1447, 1986
 A:Title: Amino acid sequence of the 203-residue fragment of the heavy chain of chicken g
 A:Reference number: A26045; MUID:87194651; PMID:3571180
 A:Accession: A26045
 A:Molecule type: protein
 A:Residues: 653-855 <ONI>
 J:Onishi, H.; Maita, T.; Matsuda, G.; Fujiwara, K.
 J. Biol. Chem. 265, 19362-19368, 1990
 A:Title: Lys-65 and Glu-168 are the residues for carbodiimide-catalyzed cross-linking be
 A:Reference number: A36604; MUID:91035476; PMID:1977747
 A:Accession: A36604
 A:Status: preliminary
 A:Molecule type: protein
 A:Residues: 54-67; 146-183 <ON2>
 R:Cole, D.G.; Yount, R.G.
 Biochemistry 31, 6186-6192, 1992
 A:Title: Stability and photochemical properties of vanadate-trapped nucleotide complexes
 A:Reference number: A43298; MUID:92329440; PMID:1385724
 A:Accession: A43298
 A:Status: preliminary
 A:Molecule type: protein
 A:Residues: 169-183 <COL>
 C:Superfamily: myosin heavy chain; myosin motor domain homology
 C:Keywords: actin binding; ATP; blocked amino end; coiled coil; hydrolase; methylated am
 F:87-777/Domain: myosin motor domain homology <MMOT>
 F:177-184/Region: nucleotide-binding motif A (P-loop)
 F:565-578/Region: actin binding #status predicted
 F:639-653/Region: actin binding #status predicted
 F:850-1940/Domain: coiled coil <COI>
 F:850-1290/Region: S2
 F:1291-1979/Region: light meromyosin
 F:1941-1979/Domain: carboxyl-terminal <CBT>
 F:2/Modified site: blocked amino end (Ser) (in mature form) #status experimental
 F:128/Modified site: N6,N6-trimethyllysine (Lys) #status experimental
 F:183/Binding site: ATP (Lys) #status predicted
 F:707,717/Active site: Cys #status predicted

Query Match 3.6%; Score 283.5; DB 1; Length 1979;
 Best Local Similarity 18.6%; Pred. No. 5.4e-05;
 Matches 252; Conservative 233; Mismatches 537; Indels 333; Gaps 54;

QY 75 LKKEFPNWEKL-VLEMHNTASLEKLVDTAWLEIMTK-----YAVGECDFFVGKPK 128
 DB 807 LARKAFKQQQLTAMKVIQRNCAAYLKLNNQWMLFTKVKPLQLQVTRBEEMQAKDEE 866
 QY 129 MLKVK-----IVKIHPLEKVDDEATEKKS-----DGACDSSDDKENSQIAQDHOKK 176
 DB 867 LQTKERQQAFAELKELEQKHTQLCEBKNLLQELQAEETELYAEEMRVRLAAKQEL 926
 QY 177 ETVVKEDGRRSINDRARR---SPRKLPSTSLKGERKWAHPKFLPHKYDVVKLQNEDKII 233
 DB 927 EELHMEARIEEERSSQQLAEKKQWQQLDLEQLLEBEAARQKLEKVTADGKI 986
 QY 234 SNVPADSLIRTPNPKIVYFIRHNALRACTGENAPWVVEDELVKKYSLPSKPSDFLL 293
 DB 987 KQMEDILLIMEDQ---NNKLT-----ERKLLERVSVD--- 1016
 QY 294 DPYKWTNLNPSFKRNTGSPDKSPKSKTNDSSLSPLNPKLWCHVHLKSLSG-----S 349
 DB 1017 -----LTTNLABEEKA-----KNUTKLNKHESMISEL-----EVLKKEEKSQRQLE 1060
 QY 350 PLKVNKNKNSKSPPEHLEEMKQWSPNKLHTNFHPKGGPAKPD-----GKHSQKPL 402
 DB 1061 KIKRKLGESSDLHQIAELQIAELQAQL-----AKKEELQAALARLEDETS 1110
 QY 403 KAKGRSGKI-----LNQKSTGNSKSPK-----GLKTPYTKMQMTLLDMA 444
 DB 1111 QKNNAKKITRELESHISQLQEDLESEKAARNAEKQKRDLSLEALKTELED--TLD-T 1167
 QY 445 KGTQKWTRAPNSGGT-----PRTSKPHKHLPPAALHLIAYYKNKDNREDKRSAL 495

DB 1168 TATQQLRAKREQEVTVLKRALSETRTHAQVQEMROKHTQAVBELTEQLEQFKRAK- 1226
 QY 496 SCVTSKTARLSSBDRARLPBELSLVQKRYELLEHKRWASMSSEEQKRYLKKREBLK 555
 DB 1227 --NLDKTKQTL-EKDNADLANEIRSLSQAKQD-VEHKKK-----KLEVLQDILQ 1271
 QY 556 KKLKEKAKEREKEMLERLEKQRYEDELTKNLPAPRLVDTPBGLPNTLPGDVAMVVE 615
 DB 1272 SKYSD--GERVTELNEKVHK-LQIEVENVT-----LLNEASKNIKUTKQVATLGS 1321
 QY 616 FLSCVSGILLPDAQVPIPAVSLMALSADKGGFLYLNRLVILLQTLQDEIAEDYGBLG 675
 DB 1322 QLQDTQELLQETRQKLNVTTKLQLEDDKNS-----LQEQLEDEVE----- 1363
 QY 676 MKLSEIPLTHSVSELVRLCIRRSDVQESGSDT-DDNKOSAAPEDNEVQDEFLEKLET 734
 DB 1364 --AKQNLERHISTLTIQLSDSKKKQLEFATVETMEEGKKKQLEIREISLTQQFEKAAS 1420
 QY 735 SEFPELTSSEKLOILTALCHILMTYSVQDHMETRQMSAELMKERL---AVLKEEND-- 789
 DB 1421 YDKLEKTKNRQQQLDLDL-----VVD-LDNQRQLVSNLEKKQKQKFDQMLAEKNIS 1470
 QY 790 KKRAEKOKRKEWAKNKEN-----GKVENGLGKTDRKKRVKFPQVDTEAEDMISA-- 841
 DB 1471 SKYADDERAABAREKETKALSARALEAL---EAKELERTNKMKAEMEDLVSSKD 1527
 QY 842 -----VKSRRLLAIQAKEIREIQREMKVKLER-----QAEFE---RIRKHAAB 884
 DB 1528 DVGKNVHELEKSKRTL-----EQVEEMKTQLELEDEBELQAAEDAKLRLVNMQAMK 1579
 QY 885 KAFQEGI-----AKAKLVWRTPPIGTDRNHNRYLWFSDEVPGLFTEKGWVHDSIDYR 936
 DB 1580 SQFERDLQARDEQNEKKRQLLKQ-----LT-----LH----- 1605
 QY 937 FNHHCKDHTVSGDEIDCPRSKKANLGKNASMNTQH-----GTATEVAVVTTTPKQOQ 988
 DB 1606 -----EHETELEDERRQALAAAANKKLEVDVLDLSQVDSANKAREEAIKQLRLQAQ 1659
 QY 989 NLWFLCDSQKELD-----ELNCLHPQGITRESQKLERLERYODIITHSLHARKPNIG 1041
 DB 1660 ---MKDYQRLDLDARAAREEIFATARENEKAKKNLEAELIQLQEDLAAERARKQADLE 1715
 QY 1042 LKSCDGNQLNLFSLDLIEVATRLQKGLGVVEETSEFEARVISLEKLDKGFGEVIALQ 1101
 DB 1716 -----KEENAEELASANSRTSLQDEKKRLREARIAQLESELDEEHENIETM 1761
 QY 1102 ASVIKKFLQGFMAPKQKRKLQSDSAKTEEVDEKKNVEBAK-VASALEKWKTAIRE-- 1158
 DB 1762 SDRMKAVQ---QAEQLNNELATERATAQKNENARQQLERQNKELRSKLOEMEGAVKSKF 1818
 QY 1159 AQTFSRMHVLLGMLDACIKWDMSAENARCKVCPKGGEDDKLILDECNKAFHLCRLPAL 1218
 DB 1819 KSTIAALEAKTASLEQLEQEAEREKQAAAKTLRQK--DKK-----LKDAL 1861
 QY 1219 YEVPDGEWQACOPATARNRSRGNYTE--ESASEDSEDDSEDEEEEEEED-- 1274
 DB 1862 LQVEDERKQABOYKQDAEKGNLRKLQKRLQLEEAEEESQRINANRRKLQRELDATESND 1921
 QY 1275 --YEVAGLRPRKRTIRGKH--SVIPPAARSGR 1304
 DB 1922 ALGREVAALKSLR---RGNEPVSFAPRRSGRR 1953

RESULT 20

A39638

Plectin - rat

C:Species: Rattus norvegicus (Norway rat)

C:date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004

C:Accession: A39638; S21876

R:Wiche, G.; Becker, B.; Lubber, K.; Weitzer, G.; Castanon, M.J.; Hauptmann, R.; Stratawa

J. Cell Biol. 114, 83-99, 1991

A:Title: Cloning and sequencing of rat plectin indicates a 466-kD polypeptide chain with

A;Reference number: A39638; MUID:91268156; PMID:2050743	
A;Accession: A39638	
A;Status: preliminary	
A;Molecule type: mRNA	
A;Residues: 1-4687 <WIC>	
A;Cross-references: UNIPROT:P30427; EMBL:X59601; NID:g1292885; PIDN:CAA42169.1; PID:g156	
C;Superfamily: plectin; alpha-actinin actin-binding domain homology; ribosomal protein S	
C;Keywords: cytoskeleton; transmembrane protein	
F;6-103/Domain: ribosomal protein S10 homology <RS10>	
F;184-399/Domain: alpha-actinin actin-binding domain homology <ACT>	
Query Match 3.5%; Score 282; DB 1; Length 4687;	
Best Local Similarity 17.7%; Pred. No. 0.00017;	
Matches 330; Conservative 280; Mismatches 660; Indels 592; Gaps 76;	
QY	35 REEYEA---RLERYSRRIWTKSTGSSQLTHKEAWEEQEVALLKEEPPA-----82
DB	795 KESYSALMELEMKKEKKEIQNTGD-----RLLEDHPARTVYESFOA 838
QY	83 -----WYEKLVLEWHHTASLEKLVDTAWLEIMTKYAVGECDFEVGKEMKLVKIVKI 137
DB	839 ALQOTW--SWMLQCCCEIAHLKE--NTAYFQFSDVREAE-----876
QY	138 HPLEKVDGEATEKSDGACDSSDKENSSQIAQDHQKXETVVYKDEGRRESINDR-- 195
DB	877 -QQLQOETLRRKYS---CDR-SITVTRLEDLLQDAQDEKEQLNEYKGHLSGLAKRAKAI 931
QY	196 -----RSP-----RKLPISLKGKRWKAPPKPLPHKYDVKLQNKDKIIS 234
DB	932 VOLKPNPAHPVRGHVPLLAVCYDKQVEVTVHKGDCQCLVGPAPPHW-----KVLS 983
QY	235 NVPADSLIRT-----ERPPNKE---IVRYFIRHNA-----LRAGTGNAPVWV---BD 276
DB	984 SSGSEAAVPSVCLVPPPNQEAQEAQVALEAQHQLVTLWHQLHVDKMSLLAQSLNRDI 1043
QY	277 ELVKYKSLSKSDFLDPKYMTINPSTPKK-----NTGSPDRK 316
DB	1044 QIRWSLVLT-----FRLKPEQQRALNRLNHLHYQAFLRDSQDAGFGPEDRL 1092
QY	317 PSKKSTDNSSLSPLNKLWCHVH---LKKLSGSPKLVKNSKNSKSPPEHLEEMKWM 373
DB	1093 VAEREYGS-----CSRHYQQLLOSLEQGEESRCQRCISBLKDIRLOLEAC 1139
QY	374 SPNKLHNFHPIPKGPPAKPGKHSKPLKAGRSKGIILNGQSKTGNKSPKGLTKPTK 433
DB	1140 ETRTVH-RLRLPLDKDPAECARIAEAOQKAAQAEVEGL--GKGVARLSAAEKVLALEP 1196
QY	434 KMKQMTL---LDWAKGTQKMTA-----PNSGCTPTSPKPKHLP-----472
DB	1197 SPAAPTLSRELELTGLKQVRSLSAIYLEKLKTIISLVIRSTQGAEEVLKTHEELKQEA 1256
QY	473 --PAALHLIAYYKENKDRSALSCVISTARLLSSDRARLPPELRSLVQKRY---526
DB	1257 AVPATLQLELVTKASLKLRAQEAQPPVNTLR-----DELGAQEVGERLQORHGERD 1311
QY	527 -----ELLEHKRW---ASMSBEQRK-----EYLKKRBEELKKKLK---559
DB	1312 VEVERWRVTLLE---RWQAVLAQTDVQRLEQLGRQLRYRESADPLSSWLQDAKS 1368
QY	560 -----EKAKE--REKEMLERLEKO-----KRYEDQELT 586
DB	1369 ROEQIOAVPIANSQAAREQLRQEKALLEIERHGEKVEBCQKFAKOYINAIKDYELQLIT 1428
QY	587 GKNLPAPRLVDTPGELPNTLFGDVAMVVEFLSC---YSGLLLPDAQYPIITAVS-----636
DB	1429 YK--AQLEPVASPAKPKVQSGSESVIQYVDLTRYSELTLTTSQY-IKFISETLRRME 1485
QY	637 ----LMEALSADGGFLYLNRLVILLQTLQDEIAEDYG-----ELGKMLSE- 680
DB	1486 EBERLAEQQRABE-----RERLAEVALEKQRLAEAAHAQAKAAQAELEARELQRMQEE 1540
QY	681 -----IPLTHSVSELVRLCLRRSDVQEESGSDTDNKGSAAPFEDNEVDQEFLE 730

myosin heavy chain, smooth muscle, long splice form - rabbit
 N:Contains: myosin ATPase (EC 3.6.4.1)
 C:Species: Oryctolagus cuniculus (domestic rabbit)
 C>Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 19-Apr-2002
 C:Accession: A41604; A33501
 R: Babji, P.; Kelly, C.; Periasamy, M.
 Proc. Natl. Acad. Sci. U.S.A. 88, 10676-10680, 1991
 A:Title: Characterization of a mammalian smooth muscle myosin heavy-chain gene: complete
 A:Reference number: A41604; MUID:92073350; PMID:19611735
 A:Accession: A41604
 A:Molecule type: mRNA
 A:Residues: 1-1972 <BAB>
 A:Cross-references: GB:M77812
 R: Nagai, R.; Kuro-o, M.; Babji, P.; Periasamy, M.
 J. Biol. Chem. 264, 9734-9737, 1989
 A:Title: Identification of two types of smooth muscle myosin heavy chain isoforms by cDN
 A:Reference number: A33501; MUID:8925535; PMID:2722872
 A:Accession: A33501
 A:Molecule type: mRNA
 A:Residues: 1455-1972 <NAG>
 A:Cross-references: GB:J04833; NID:gl65519; PIDN:AAA31407.1; PID:gl65520
 A:Experimental source: smooth muscle
 A>Note: examination by Southern blotting for the regions of difference between this isofo
 active splicing
 C:Superfamily: myosin heavy chain; myosin motor domain homology
 C:Keywords: actin binding; alternative splicing; ATP; coiled coil; hydrolase; methylated
 F:88-771/Domain: myosin motor domain homology <MMOT>
 F:178-185/Region: nucleotide-binding motif A (P-loop)
 F:559-572/Region: actin binding #status predicted
 F:633-647/Region: actin binding #status predicted
 F:844-1938/Domain: coiled coil #status predicted <COI>
 F:844-1284/Region: S2
 F:1288-1972/Region: light meromyosin
 F:1939-1972/Domain: carboxyl-terminal <CBT>
 F:129/Modified site: N6, N6-trimethyllysine (Lys) #status predicted
 F:184/Binding site: ATP (Lys) #status predicted
 F:701,711/Active site: Cys #status predicted
 Query Match 3.5%; Score 281.5; DB 1; Length 1972;
 Best Local Similarity 19.2%; Pred. No. 6.6e-05;
 Matches 266; Conservative 219; Mismatches 503; Indels 399; Gaps 60;

QY	75	LLKEEPPANVEKL-VLEMHNTASLEKLVDTAWLEIMTK-----YAVGECDPFEVGEK	128
DB	801	LARKFAKQQQTAKTKVIRQNCAYLKNQWMLFTKVPQLQVTRQEEEMQAKDE	860
QY	129	MLKVK-----IVKIHPLEKVDDEATEKKS-----DGACDSPSSDKENSSQIAQHQKK	176
DB	861	LQIKERQQAESLQELQKHTQLSEKNLLQEQQAETELYAAEEMRVRLAAKQEL	920
QY	177	ETVVKEDGRRSINDRARRSPKLPSTLKKGERKAPPKFLPHKYDVVKLQNEDKIISNV	236
DB	921	EEILHEMEARLEBEDRGOQ-----LQAEK-----KMAQQMLDLBEQLEEE-----	961
QY	237	PADSLRTERPPNKEIVFIRHNALRACGENAPVWVEDELVKKYSLPKES-----DFLL	293
DB	962	-----EEAARQKLQ-----LEKVTAEAKIKKLEDDILVMDQNNKLKSKERKLLIE	1005
QY	294	DPYKVTLPNTPSKTKNKGSPDRPKSKSTDNSSLSPLNKLWCHVHLKKSLSG-----S	349
DB	1006	ERISDLITNLABEEKA-----KNLTKNKHESMISEL-----EVRLKKEKSQLE	1054
QY	350	PLKRVKNSKSGPPEEHLKEEMKMSPNKLHTFNPHPKGPPAKPP-----GKHSKPL	402
DB	1055	KLKRKMDGEASDLHEQIADLQAIKLMQ-----AKKEELQAALARLEDETS	1104
QY	403	KAKGRSKGI-----LNGQKSTGNSPKK-----GLKTPTKMKQMTLLDWA	444
DB	1105	QKNNAUKLIRELEGHISDLQEDLSRRAARNKAOKRDLGEELEALKTELED--TLD-T	1161
QY	445	KGTQKQTRAPRNSGGT-----PRTSS-----KPHKHLPPAALHLIAYKENDKOR	488
DB	1162	TATQQLRAKREQEVTVLKALDEETRSHEAQVQEMRQKHTQVVEELTQLEQFQKAKAN	1221

RESULT 22

T18296

myosin heavy chain - Entamoeba histolytica

C:Species: Entamoeba histolytica

C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004

C:Accession: T18296

R:Guillen, N.

Db 177 EGDKORT--KKGTKGSGGNKGRSTKE--NRAKKG-----KLVKURFTQKLEHCIK 226
Qy 229 -----EDKIISNVADSLIRTERPPNKEIVYFIRHNALRAGTCENAPWVE 275
Db 227 SEWADVCECRNFTEDEKRLAASYKCMGTETK-----IKSICRKRVI-AEMCEAAGYVKS 280
Qy 276 DELVVKYSLPFSDFLLDPYKVTMLNPSTKNTGSPDRKPSKSKTNSLSPLNPK 335
Db 281 SEPKKK-----GKKKK-----DEKKEBELERILKEQABEAK 314
Qy 336 LCHVHLKSLSGSLPKVNSKN-----SKSPEEHLLENMOMSPNK 377
Db 315 IRGVVVKKKCEKALK-KKCKDLGRKMKEEAKKCAALAKKQKEDEKACKELAKK 373
Qy 378 LHTNHIPKGGPPAKPGHSDKPLKA--KGRSKGILNGQKSTGNSKSPKGLKTPKTKM 435
Db 374 -----KEADEKKCEEAANKKAAEKKEKCEAAEKKEKCEAAEKKEA 425
Qy 436 KOMTLLDMAKGTQKMTAPRNSGGTPTSSKPKHLPAPALHLIAVYKENKDRED--KRS 493
Db 426 ERKKCEELAKNIKAAEKKKCEAAKKEBAERKKCEELAKKI KAAEKKKCEETAKG 485
Qy 494 ALSVISKTARLLSSSDRALPPELRSLVQRYELLEHK-----KRWASMSSEORKEYL 547
Db 486 KEVAERKKCEELAKKI KAAEKKKCKLAKKEKETAEEKKCEAAKKEKKEAEEKKCEKA 545
Qy 548 KKRREL--KKLKEKAKERRE-----KEMLERLEKQ-----RYE 581
Db 546 AKRKEAAEKKEKCEKSAKRRKEAAEKKKCEKAAKERKEAAEKKKCEEAKEKVAERKK 605
Qy 582 DOELTGKLPAPRLVDTPEGLPNTLPGDVAMVVEFLSCVSGLLPDAQVPTAVSLMEAL 641
Db 606 CEELAKK-----IKKAAEKKKCKE-----AAKKEKAA 633
Qy 642 SADKGSFLYNRLVILLQTLQDEIAEDYBELG-----MKLSEIPLTHLSYSELVRLCL 696
Db 634 EREKCG-----ELAKKI KAAEKKKCKLAKKEKETAEEKKCEKAAK 675
Qy 697 RSDVQESSEGGSDTDNDKSAFED--NEVQDEPLEKLTSEFFELTSEKQIILTALCH 754
Db 676 KRKEAAEKKKCEAAKKEKAAEKKKCEAAKKEKAAEKKECEELAKKIKKAAEKKKCK 735
Qy 755 RILMTYSVODHMETR-----QOMSABLWKERLA-----782
Db 736 KLAKKKAGEKVKLKGNGKGGKALKKKKKCEELAKKGAEEKKCEAAKKEKAAEKKK 795
Qy 783 -----VLKEENDKKRAEK--QKRKEM-----EAKNKENGKRYENGLGKTD 819
Db 796 CEKTAKKRKEAEKKCKCENTAKKRKEAEKKKCEAAKKEKKEAEKKCKCEKAKKETA 855
Qy 820 RKKRIVKFPQVDTEAE--DMISAVKSRLLAIQ-----AKEREIQREM-----863
Db 856 EKKKCEKAAKGRKQAEKKKCEKAAKGRKEAAEKKKCEAAKKEKELAEKKCEBAAKKE 915
Qy 864 KVKLERQAEERIRKHGAAAEKAFQBGIAKAKLVNRRTPIGTDNRHNRVWLFSDVPGLF 923
Db 916 KEVAERKKCEELAKKI KAAEK-----KKCKLAKKEKAGEKN-----954
Qy 924 IEGWVHDSIDYRNFHHCKDHTVSGDEDYCPRSKKNANLGNASMTQHGTATEVA---VE 980
Db 955 -----KLKKKAGK-----KKCKKLGKKKRAAEKKKCAEAKEKE 992
Qy 981 TTTPKQGNLWFLCD-----SQELDELNLCHPHQIGRESQLEKRYQDIIHSIHLAR 1036
Db 993 AATKKK-----CEERAKKQEAEEKQCEE-----RAKKLKEAAEQ-----1028
Qy 1037 KENLGKSGDGOELNLFU--RSDLLEVATRLQKGGVVEETSPFEARVISLEKLDKPG 1094
Db 1029 -----KQCEERAKKLKEAAEKKKQCEERAKKLKEAA-----EQKQCEERAKKLKEAAEKK 1077
Qy 1095 ECVIAL-----QASVTKKFLQGFMAPKQKRKLQSEDSAKTEVEDEKGVBEA-----1143
Db 1078 QCEERAKKEAAEKKQCEERAKKLKEAAEKKKQCEERAKKEAAEKKKCEEAKEKEA 1137

Qy 1144 ----KVASALEKWKTAIREAQTFSRMHVLLGMLDACIKWDMSAENARC-----1187
Db 1138 AEKKKCAEAAKKEKEATEKQK-----CAEAAKKEKEAAEKKKCAEAARKKBAQA 1187
Qy 1188 -KVC-----PKKGEDDKLILCDECNK-----AFHLFCLRPALYVEVDPGWCQCPACQATAR 1237
Db 1188 KKKCADLAKKEQEPAEKMKCEBAEKKEAAEKKQKCAAAKKEKAAEKK--KCAEAARK 1245
Qy 1238 RNSGRNVTESASDSDSDSDSEDEEBEEDYEVAGLRUR--PKTIROKHSHVI 1295
Db 1246 -----EQEAAEKKKCAEAAKKEKKEKKEKKEKAAALKRQCAKLVIRAKEAAL 1296
Qy 1296 PP-----AARSRRPGKPHSTERSQPKAPPVDDDAVDELVLQTKRSSROSLEOKCEBI 1351
Db 1297 RKCKAILAKKAMAAEKKECEKLAJ-KA-----KEAIEWKCAKLAKKRREAKKKCAKL 1350
Qy 1352 LHK 1354
Db 1351 AKK 1353
RESULT 24
S28261
centromere protein E - human
N:Alternate names: centromere 312K protein; kinesin-related protein CENP-E
C:Species: Homo sapiens (man)
C:Date: 03-Mar-1994 #sequence_revision 03-Mar-1994 #text_change 09-Jul-2004
C:Accession: S28261
R:Yen, T.J.; Li, G.; Schaar, B.T.; Szilak, I.; Cleveland, D.W.
Nature 359, 536-539, 1992
A:Title: CENP-E is a putative kinetochore motor that accumulates just before mitosis.
A:Reference number: S28261; MUID:93024922; PMID:1406971
A:Accession: S28261
A:Molecule type: mRNA
A:Residues: 1-2663 <YEN>
A:Cross-references: UNIPROT:Q02224; EMBL:Z15005; NID:g29864; PIDN:CAA78727.1; PID:g29865
C:Genetics:
A:Gene: GDB:CENPE
A:Cross-references: GDB:361164; OMIM:117143
A:Map position: 4q24-4q25
C:Superfamily: centromere protein E; kinesin motor domain homology
C:Keywords: ATP; coiled coil; microtubule binding; mitosis; nucleotide binding; P-loop
F:7-335/Domain: kinesin motor domain homology <KMT>
F:86-93/Region: nucleotide-binding motif A (P-loop)
F:486-2183/Domain: coiled coil #status predicted <CoI>
F:92/Binding site: ATP (Lys) #status predicted
Query Match 3.5%; Score 280.5; DB 1; Length 2663;
Best Local Similarity 19.3%; Pred. No. 0.0001;
Matches 337; Conservative 302; Mismatches 598; Indels 509; Gaps 82;
Qy 39 EARLERYSERIWTCKSTGSSQLTHKEAWEEBOEVAELLKEFPFAYWEKLVLEMVHHTAS 98
Db 788 ESRVQGLLEEIGTKDLDLATTQSNYKSTDQEFQNFKLHMDPEQKY-KMVL- --ENERM 843
Qy 99 LEKLVDTAWLEIMTKYAVGECDFEVGKEKMLVKVI-KVIHPLEKVDSEATEKXSDGACD 157
Db 844 NQEIYN-----LSKEA--QKPDLSLG---ALKTELSYKTOELOEKTREVQERLNE----888
Qy 158 SPSSDKENSSQIAQHQKKEKTVVKEDEGRRESINDRARRSPRKLPT-SLKGERKWAPPK 216
Db 889 -----MEQLKEQLENRDSPLQTVREKTLITEKLOQTLTEEVKTLTQEKDDLQQLQBS 940
Qy 217 FLPHKYDVKLQNEDKIISNVPADSLIR-----TERPPNKEIVYFIRHNALRAG 265
Db 941 LQIERDQLKSDIHTDVNNIDTQEQLRNALESQKQOETINTLKSISSEVSRNLHMBEN 1000
Qy 266 TGENAPVVEDELVKKYSLPKSFSDFLDDPKYMTNLNSTKCKNTGSPDRKPSKSKTND 325
Db 1001 TGE-----TKQBFQOKMGVIDKKQD--LEAKNTQTLTADVKNDEIIEQORKIFSLIQEKN 1053
Qy 326 ---SSLSSPLNPKLVCHVHLKSKSLSGSPLKVKNSKNKSPBEHLEEMKMMSPNKLHTNF 382

Db 1149 DAIBETNEKKAARQAKARRDMAELESYKQ--ELEESNDKTVLHSQ-----LKAKR 1199
Qy 552 EE-----LKKLKEKAKERREKEMLERLEKQKRYEDELTKGNLPAPRLVDTPEGHPNTL 607
Db 1200 DEEYAHLOQLEETVKS--SEEVBEKMAQNOKTIEELN-----ETIDQL----- 1242
Qy 608 GDVAMVVEFLSCYGLLLPDAQYPIITAVSLMEALSADKGGF--LYLNRVLVILLQTLQDE 666
Db 1243 -----KKQKISADKAKASASSNDNFRAELSNJASARLEAKKRRK 1282
Qy 667 IAE-DYGEIGMKLSBIPILTHSVSLVRLCLRSVDQVESEGSDDDNKDSAAFDEYQ 725
Db 1283 AAETSLMEKDKWREMQSNLDL--MAKLSKNNELESIQKASADETLNSLLKKNASL 1340
Qy 726 DFLEKLETSSEFELTSEKQLIQLALCHRI-----LMTYSVDHMETRQMSAE-- 775
Db 1341 D-----MQLSELTE-ASEEDRTRATLNKIRQLEEDLAVAVEARDALDAQETKEVK 1394
Qy 776 ----LWKERLAVLKEENDKKRAE--KQEKEMAKNKGKVENGLGKTDR--KKRIVKPE 828
Db 1395 EVKSLAEARKKLDENREVMELRKKKEKELUSAKERADMAEQARDKAAERAKKAIQEA 1454
Qy 829 POVDTAEADMI SAVSRRLLAIQAKKERIQEREMKVKLERQAEER-----IRKHAAA 883
Db 1455 EDVQKELTDVAATRE-----MERKMR-KFDQQLAEERNNTLLAQQRDWA 1499
Qy 884 EKAFQEGIAKAKLVWRRTPIGTDRNHNRYWLFSDVPGLFIEKGWVHDSIDYRFNHCKD 943
Db 1500 HQMLRDAETKA-LVL-----SNELS-----EKQDIVDQLE-----KD 1530
Qy 944 HTVSGDEDYCPRSKCANLGNKASMNTHGCTATEVAVETTPKQGNLWFLCDSQKELDEL 1003
Db 1531 -----KRTLKLEIDNLAS-----TKDAGKNVYELETKRRLDDE 1565
Qy 1004 LNLCHPQGIT-----ESOLKREKRYQDIIHSIHARKPNLG 1041
Db 1566 LSRAEQIILEDALQADARSVENVNQMRSFERQLASREED-----EDDRKKGIT 1620
Qy 1042 LKSCDGNQBLNPLR-----SDLE-----VATRLQKGLGV 1074
Db 1621 SKIRNLTESEFQARQAAIANKKIESQISELTEKNEASLRQIEDLSRQLRKAQLGKW 1680
Qy 1075 E-ETSEFEARVLSLEKLDPGECVIALQAS--VIKKFLOGFWAPKQKRLKQSDSAKTE 1131
Db 1681 DIQLDVTAEARAMEDALQRAEKRAKASEDEIKRLTIADIOAVSSSKRAEAREDELIE 1740
Qy 1132 EV-----DEEKWVBEAKV-----ASALEKWKTAIREAQTFSRMHVLGML 1172
Db 1741 EVSSLRASSFSNEEKERL-EAKVIDLEDQDEASANELAQEKVRKQQ-----QLEGM 1793
Qy 1173 DACIKWMSAENARCKVCPKQGEDDKLILCDENKAFHFCURPALYEVDPDGEWQCPAQ 1232
Db 1794 TA-----DLAMERSVC-----ERTESOKIAL-ERAN-----RDLKQQLQDAE-----N 1830
Qy 1233 PATARNRGRNVYTESASEDSEDESEDESEDESEDESEDESEDESEDESEDESEDESE 1292
Db 1831 TAVARLURTO-INVAEKVS-----SLEQQLSLEEQDKMQGRTLRMRET---KM 1875
Qy 1293 SVIPPAARSGRPGKK-----PHSTRSQPKAPPVDDAEVDELVLQTK-RSSRRQSLEIQ 1346
Db 1876 AEMQOMLEBEKQGESNRQAVDRQNARIRQLRTQLEDTEARDRLTNKLDERRAEEMT 1935
Qy 1347 KEEELHK---IVKTRFSPFPREP 1367
Db 1936 DLNETLSRDVSLKQRETTARRTP 1959

RESULT 28

A36014

myosin heavy chain, nonmuscle - fruit fly (Drosophila melanogaster)

N:Contains: myosin ATPase (EC 3.6.4.1)

C:Species: Drosophila melanogaster

C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 09-Jul-2004

C:Accession: A36014; B36014
R:Ketchum, A.S.; Stewart, C.T.; Stewart, M.; Kiehart, D.P.
Proc. Natl. Acad. Sci. U.S.A. 87, 6316-6320, 1990
A:Title: Complete sequence of the Drosophila nonmuscle myosin heavy-chain transcript: c-
A:Reference number: A36014; MUID:90349606; PMID:2117279
A:Accession: A36014
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-2017 <KET>
A:Cross-references: UNIPROT:Q99323; GB:M35012
C:Genetics:
A:Gene: FlyBase:zip
A:Cross-references: FlyBase:FBgn0005634
C:Superfamily: myosin heavy chain; myosin motor domain homology
C:Keywords: actin binding; alternative splicing; Atr; coiled coil; hydrolase; methylated
F:1-2017/Product: myosin heavy chain, form I #status predicted <MAI>
F:46-2017/Product: myosin heavy chain, form II #status predicted <MA2>
F:135-815/Domain: myosin motor domain homology <MMOT>
F:225-232/Region: nucleotide-binding motif A (P-loop)
F:598-631/Region: actin binding #status predicted
F:703-727/Region: actin binding #status predicted
F:888-2017/Domain: coiled coil #status predicted <COI>
F:888-1328/Region: S2
F:1329-2017/Region: light meromyosin
F:176/Modified site: N6,N6-trimethyllysine (Lys) #status predicted
F:231/Binding site: ATP (Lys) #status predicted
F:745,755/Active site: Cys #status predicted

Query Match 3.5%; Score 277; DB 1; Length 2017;

Best Local Similarity 19.9%; Pred. No. 0.00011;

Matches 286; Conservative 223; Mismatches 483; Indels 446; Gaps 66;

Qy 60 LTHKEAMEEORVAEALLKEEPPAW-----YEKLV-----LEMVHNTASLEKLVDT 105
Db 819 LAHLEE-ERDFKISDLI-VNFQAFGRFLARRNYQRLQQLNAIRIIORNCAYLKLNRW 876
Qy 106 AWLEIMTK-----YAVGECDSEVGEKMKLVKVIKHPLEKVDDEATEKSDGACDSP-- 159
Db 877 QWRLYTKVPLLEVTKQEEKLVQKEDLQVREKLDLAKNTQYERYKQALVEKTYL 936
Qy 160 -----SSDKENSQIAQDHQKKTETVVVKEDEGRRESINDRARRSPRLPTSLKKG 208
Db 937 AEQLQABTELCAEAESRSRLMARKQELDMWQLETRIIEEER-----VLALGG 987
Qy 209 ERKAPPKFLPHKYDVKLQN-EDKII SNVPADSLIRTPPNKEIVRIFIRNALRAGTG 267
Db 988 EKK-----KLELNIQDLEEQLEEEEAARQKLQLEK----- 1017
Qy 268 ENAPVWVEDELVKKYSLPSPKPSDFLLDPVKYMTLPSTKRKNTGSPDRKPSKSKTDNS 327
Db 1018 -----VQJDAKIKYE-----EDLAL-----TDDNQKLLKEKL----- 1047
Qy 328 LSSPLNPKLWCHVHLKSLGSPKVKVNSKNSKSPPE-----HLEEMKMSPNKLTNPH 383
Db 1048 LEERAN-----DLSQTLAEBEEKAKHLAKKAEATITEERL----- 1087
Qy 384 IPKGPAPKPKGKSHDKPLKAG-RSKGILNGOKSTGNSKPKGLKTPKTOKMQLTLD 442
Db 1088 -----HKDQQQRESDRSK-----RKIETEVDLKEQLNERRVQVDEM-Q 1128
Qy 443 MAKG-----TQKTRAPRNSGGTPRTSSKPHKHLPPAALHLIAYYKENDREPKRSALSCV 498
Db 1129 LAKREELTQTLRRIDEES-ATKATAQAKARELESQ-----LAEIQEDLEAE----- 1174
Qy 499 ISKTARLLSSEDRARLPBELSLVQKRYELLEHKHKKRWASMSGEORKEYLKKREE-----L 554
Db 1175 --KAARAKAEKVRDLSELEAL---KNELLD-----SLDTTAAQQLRSKREQELATL 1223
Qy 555 KKLKEKAKERREKEMLERLEKQKRYEDELTKGNLPAPRLVDTPEGHPNTLFGDVAMV 614
Db 1224 KKSLEETVNHGV-----LADMRHKHQSLENSIN-----DQLENLRK-----AKTVL 1266
Qy 615 EFLSCYGLLLPDAQYPIITAVSLMEALSADKGGFLYLNRLVILLQTLQD----- 665

```
Db 1267 E-----KAGTLEAENAD-----LATELRSVNSRQEND 1295
Qy 666 -----ETAEADYGEKMLSEIPLTHSVSELVRLRRSDVQESSEG-----SDTDDNKDSA 717
Db 1296 RRRKQAESQAELOVKLAET-----RARSELOEKTCK-----LQQAENITNQLAEAEIKASA 1349
Qy 718 AFEDNEVDQFLEKLETSFELTSEBKQILFALCHRIILMTYSVQDHMETROQMSAELW 777
Db 1350 AVKSANMESQL--TEAQQLLEBETROKGLSLK-----RQIESE-- 1388
Qy 778 KERLAVLKEEND-----KKRAEKQK--PKMEAKNKENGK----- 810
Db 1389 KEALQOLEDEDDAKNYERKLAETVTTQOEIKKAEEDADLAKELEEGKRLNKDIEAL 1448
Qy 811 ---VENGLGKTRDKRIGVPEPOVTEAEDMISAVKSRRLLAIAQAKEREIQREMKVKL 867
Db 1449 ERQVKELIAND---RLDKSKKKIQSELEDAITIEEAQRTKVLEKK---QKQFPKILA 1502
Qy 868 EROAEEIRKHKAAAEKAFQEGIAKAKLVMRTPITGTR---NHNRYWLFSDVFPGLFT 924
Db 1503 EERAISEQIAQERDTAEREAREKETKLVSVSRLEDAFDKIEDLENKRTKLQNELODLAN 1562
Qy 925 EKGWVHDSIDYRFNHHCCKOHTVSGDEDEYCPRSKANKLGKVASMTQHGTAETAVETTP 984
Db 1563 TOGTADKNV-----HELE-----KAKRALESQALAEKLAQNEELED----- 1597
Qy 985 KQCONLWFLCDSQKELDELINCLHPQIRESQKLER-----LEKRYODIILHSLAR 1036
Db 1598 ---DLQTEDAKLRLEVNMQALRSQFERDLAKEBAGEKRGVLVQJRDLETELEDEER 1653
Qy 1037 KPNLG-----LKSCDGNQELLNFLRSLIEVATRLQ---KGGLGYVBETSEFEA 1082
Db 1654 KORTAAVASKKLEGLDKETIETTMEMHNKVKEDALKHAKKQAOVQDALARDAEEAKAKE 1713
Qy 1083 RVLSLEKLDGFCVIALQASVTKKFLQGFMAKQKRGKLOESDSAKTEE----- 1132
Db 1714 ELQALSKEAD-GK-VKALEAVL-QLTEDLASSERARRAAETEDRLAEFEIANNANKSL 1770
Qy 1133 -VDEEKMYVEEAKVASALEKWKTAIREAOTFSRMVLLGLMDACIKWMSAENARCKVCP 1191
Db 1771 MIDEKERL--EARIATLEEB---LEEQSNSEV-----LLDRAARQIQIEQLITELAN 1819
Qy 1192 KXGEDDKLILDCENKAFHLFCLRLPALYEPDGEWCQACOPATARNRGRNYTESSAS 1251
Db 1820 EKSNQK---NENGR-----ALLERQNKELKAKLAETIAQK-----TKVKAT 1859
Qy 1252 ESEDEDESDEEEEEEEDYEVAGLRLRPRTIIRGKHSVPPAARSGRPPGKK-PH 1310
Db 1860 IATLEAKIAKVEQLENEGKE-----RLQOK-----ANRMDKKIKE 1897
Qy 1311 STRRSQKAPPVDD--AEVDEL---VLQTKRSTRRSQSLQK-----CEEIL 1352
Db 1898 LTNWIEDRRHVQHKQEQMDKLSRIKLLKRNLDTEEBLQKEKTKQKRYQRECEDMI 1955

RESULT 29
A59404
plectin [imported] - human
C:Species: Homo sapiens (man)
C:Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 09-Jul-2004
C:Accession: C59404; A59404
R:Liu, C.G.; Maercker, C.; Castanon, M.J.; Hauptmann, R.; Mische, G.
Proc. Natl. Acad. Sci. U.S.A. 93, 4278-4283, 1996
A:Title: Human plectin: organization of the gene, sequence analysis, and chromosome loca
A:Reference number: C59404; MUID:96210632; PMID:8633055
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-4684 <STO>
A:Cross-references: UNIPROT:Q15149; GB:CAA91196; NID:91296662; PIDN:CAA91196.1
C:Superfamily: plectin; alpha-actinin actin-binding domain homology; ribosomal protein S
```

```
Query Match 3.5%; Score 277; DB 2; Length 4684;
Best Local Similarity 17.8%; Pred. No. 0.00029;
Matches 278; Conservative 242; Mismatches 556; Indels 482; Gaps 57;

Qy 25 IPIHQEAFPTREYE--ARLERYSERIWTCKSTGGSQLTHKE-AWEEEOEVAEL----- 75
Db 1471 IKFISLTLRRMBEERLEAQRAER-----ERLAEVAALBKQRLABAHQAQA 1521
Qy 76 -LKEFPFANYEKLVLMEVHHNTAS-----LEKLVDTAWLEITMTKYAVGEECD 121
Db 1522 QAEREAKELQRIQIEEVVREANAQVDAQQKBSIQEELQOLROSSAEATQAKARQAEAE 1581
Qy 122 ----FEVGEKMLKVKIVKIHPLEKVDDEATEKKGDCGCDSPSDKENSQIAQDHOKKET 178
Db 1582 RSLRLRIEEB---IRVRL-----QLEATERQGGG----- 1608
Qy 179 VYKEDGREGRESINDRARRSPKPLTSLKKGKRWAPPKPLPHYKYDVKLQNEDKIISNVPA 238
Db 1609 ----EGELOALRARAEAEAKQKQAEAEAL-----RRQVDESQRKQAEVEL 1654
Qy 239 DSLIRTERPPNKEIVRYFIRHNALRAGTCENAPVWVVEDELVKYSLSPKSFDFLLDPYKY 298
Db 1655 ASRVAAEAAAEKQALQALAEELRLQAEAEERWLQAEVERARQV-----Q 1701
Qy 299 MTLPSTTRKNTGSPDRKPSKSKTDNSSLSPNPKLWCHVHLKSLSGSPLKV----- 353
Db 1702 VALETAQRSAAEALQSKRASFAEKT-----AQERSLOEHHVAVAQLE 1745
Qy 354 ---KSKNSKSPBEHLEEMWOMSPNKLHTNFHIPPKGPPAKPKPGKHSKDKPLKAKGRSKG 410
Db 1746 EAERRAQQAQAEAREAEAEERQERWLKAN-----EALRLRLQAEAEVLQOKSLAQ 1797
Qy 411 ILNGKSGTSGNSKPKGLKTPKTKMKQMTLLD-----MAKGTOMKTRAPNSGGTTPR 462
Db 1798 EAEKQKEAEERARRRG-KAEQAVRQRELAQOELEKQRLAEGTAQORLA----- 1847
Qy 463 TSSKPHKHLPPAALHLIAYYKENKDREKRSALSCVISTARLLSSEDRARLPEELRSLV 522
Db 1848 -----AEQELIRLRAETEQEQEQORQL-----EEELARLQREAAAAAT 1884
Qy 523 QKRYEL-LEHKKRWASMBEEOKEVYL---KKRBEELKKLKBKAKERRKE----- 569
Db 1885 QKQELEAEALAKVRAEM-----EVLASKAKAEESRSTSEKSKORLEBAEAGRPRELAE 1938
Qy 570 -----MLERLEKOKRYEDQELTGKPLPAFLVDTPEGLPNTLFGDVAMVVEFLSCYSG 622
Db 1939 EAARLRLAAEAKRQRLAEEDAAQRABAEV----- 1971
Qy 623 LLLPDAQYPTITAVSLMEALSADKGGFLVNLRVLVILLQTLLODE-----IAEDYGEKMLK 678
Db 1972 -----LAEKLAIGEATRLKTEAEIALKEKEAENERLRLRAEAEAFORREL 2017
Qy 679 SEIPLTLHVSSELVRLCLRRSDVQEE---SGSDTDDNKDSAAPE-----NEVODEFL 729
Db 2018 EE-----QAAQHKADIEERLAQLRKASDSELERQKGLVEDTLRQRQVBEETL 2065
Qy 730 EKLETSEPFELTSEBKQILFALCHRIILMTYSVQDHMETROQMSAELWKLAVLKEEND 789
Db 2066 -ALKAS---FEKAAAGKAELELELGR---IRSNAEOTLRSKEQAELEAAKQRLAEERER 2119
Qy 790 KKRAE-----KQKRKEMEAKNKENGKVENGLGKTRDKRKRIVKPEQPVDTE- 834
Db 2120 REEAERVQKSLAAEAEAAARQKAALEEVERLAKVVEEARSRLRAEQESARQLQAQEA 2179
Qy 835 AEDMISAVKSRRLLAIAQAKEREIQE-----REMKVKULE 868
Db 2180 AQKRLQAEKKAHAFVQ-QKEQELQOTLQEQESVLDRLRGEAAAEARRAAEEAEARVQAE 2238
Qy 869 ROA-----EEERIRKH-----KAAAEK---AFQEGI---AKAKLVMRRTPI 903
Db 2239 REAASRRQVVEAEERLQKSAEAEQAAQAAQAAAEKLRKAEQEAEARRAQAAALRQKQ 2298
Qy 904 GTDRNHNRYWLFSDVFPGLFIEKGWVHDSI-DYRFNHHCCKOHTVSGDEDECYCPRSKANIG 962
```



```

Db      2299  ADAEMKHKKFAQT---LRQAQVEQLTTLRLQLBEETHQKXLLDEELQRLK----- 2350
Qy      963   KNAMNTQHGTAATEVAVETTTPKQGNLWFLCDSQKELDELNCLHPQGIRESQKLERLE 1022
Db      2351  -----AEATEAARQ-----RSQVEELFSVRVQMEELSKLKARIE 2385
Qy      1023  KRYQDIHHLARKNLGLKSCDGNQELL-----NF 1054
Db      2386  AENRALI-----LRDKNTQRFQBEAEKMKQVABEAARLSVAQAQAARLQRL 2433
Qy      1055  LRSDLI-----VATRLQKGLGVVETSEPEARVISLEKLDKDFG-ECVIALQ-----ASV 1104
Db      2434  AEDLAQRALAEKMLKEMQAVQEARLKAERLLOQKELAQARLQEDKQMAQQ 2493
Qy      1105  IKKFLQGFVAPKQKRLQSDSAKTE-----EYDEEKQMVVEAKVASA 1148
Db      2494  LAETQGFORTLEAERQRLMSAEARLKLRAVMSRAQARAEEAQRFRKAQAEIGEK 2553
Qy      1149  LEKWTATREAOQTFSRMHVLLGMLDACIKWMSAENARCKVCPKGGEDDKLILCDECNKA 1208
Db      2554  LHRTELATQKVTLVQTLRIQR-----QSDHDAERLREAIAELEREKEKL-----QOE 2602
Qy      1209  FHLFCLRPALYVVPDGEWCQACQAPATARRNS-----RGRNYTESASEDS--ED 1256
Db      2603  AKLLQLKSEMQTVQOE-----QLLQETALQOSFLSEKDSLLQRFPIEQEKAKLSQLFQD 2659
Qy      1257  DESDEEEEBEEDYEYAGRLRPRKTIRGKHSVIPPAAARGRRPKKPHSTRRSQ 1316
Db      2660  EVAKAQLREEQOQQOQQOEQRQL-----VASMEEAARRQHEAEQVRRKQ 2707
Qy      1317  PKAPPVDD--AEVDELVLQTKSSRRQSLQKCEILHKIVKVPSPFPFRPVTRDE 1372
Db      2708  BELQQLQOORQOQELLAENQRLREQ--LQLEE-----QHRAALAHSEVTSASQ 2756

RESULT 30
G64585
cag pathogenicity island protein cag7 - Helicobacter pylori (strain 26695)
C:Species: Helicobacter pylori
C:Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 09-Jul-2004
C:Accession: G64585
R:Tomb, J.F.; White, O.; Kervilavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.;
Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodde, A.; McKenne
son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Matthey, L.
Nature 388, 539-547, 1997
A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C.
A:Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.
A:Reference number: A64520, MUID:97394467, PMID:9252185
A:Accession: G64585
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1927 <TOM>
A:Cross-references: UNIPROT:O25262; GB:AE000567; GB:AE000511; NID:g2313641; PIDN:AAD0758

Query Match      3.5%; Score 276; DB 2; Length 1927;
Best Local Similarity 18.8%; Pred. No. 0.00011;
Matches 313; Conservative 237; Mismatches 580; Indels 534; Gaps 67;

Qy      139  PLEKVDDEATE-----KKSDGACDSPS-----SDKENSQIAODHOKKTV 179
Db      19   PQDLSNEEATEAHFNILKESKSSDHLNPTQTQTHFDGDKSEETQMDSEGNET- 77
Qy      180  VKEDEGRESINDRARRSPKLTSLK----- 206
Db      78   ---SESSNGSLADLPFKKARKLVNKKPFTQOKNLDDEETQELNEEDQDNEVQEBETQTD 134
Qy      207  -----KGRKAPP-----KFLPHYDKVLQWEDKIINVPADSLI 242
Db      135  LIDDETSKKTQOHSPODLSNEEATEAHFNILKESKSSDHLNPTQTQTHFDGDKSE 194
Qy      243  RTERPPNKIVIRYFIRHNALRAGTGENAPWVVEDELVKVYSLPSKPSDFLLDPYKMTLN 302

```

```

Db      195  ETQTQMDSE-----GNETSESSNGSLADLPFKK-----ARKLVNKK- 231
Qy      303  PSTKRNKGTSPDRKPSKSKTDNSLSPLNPKLWCHVHLKSLSGSLPKVKNKSNKSP 362
Db      232  PFTQOKNLDDEETQELNEEDQDNEVQEBETQTDLIDDETSKKTQOHSPODLSNEEATEA- 290
Qy      363  EEHLEEMMK-----MMSPNKLTHTNPHIIPKGPAPKPKGKSHDKPLKAKGRSKGILN 413
Db      291  -NHFENLLKESKSSDHLNPTQTQTHFDGDKS-----EITDDSDNDQBIILKSGSKKYLIIG 346
Qy      414  G-----OKSTGNSKSPKGLKTPKTKMKQMTLLDMAKGTQKQWTR--APRNSGG 459
Db      347  GIVWAVLVIILFSRSIFHYFMPLEDKSRFSKORNLYVNDEIQIROEYNRLLKERNK 406
Qy      460  TPTSSKPKHLPPAALHLIAYYKENKDRKRSALSCV----- 498
Db      407  NMIDKNLFPNDPNTRLYNLIABIEDKNPLRAFECISNGNYECLKLIKOKKLQDQ 466
Qy      499  ISKTARLL-----SSEDRARLPEELRSLVQRYELLEHKRWAAAA--SMSEORQK 544
Db      467  MKKTLEAYNDICIKNAKTEERIKCLDLIKDENLKKSLNQKQVQVALDCLKNAKTDEERN 526
Qy      545  EYLK-----KKREELKKKLE-----KAKERRE-----KEMLERLEKQ- 577
Db      527  ECLKLINDPEIREKFRKELEKQLEQYKDCIKNAKTEAEKKNKCLKGLSKEAIERLKQQA 586
Qy      578  ---KRYDEOELTGKVLPAFLVDTPEGLPNTLFGDVAMVVEFLSCYS----- 621
Db      587  LDCLKNAKTDEBERNECLKNI-----PODLQKELLADMS--VKAYKDCVSKARNEKEQ 637
Qy      622  ---GLLLPDAQVPIITAVSLMEALSADKGGFLYLNRLVILLQTLQDIAEDYDGLGML 678
Db      638  ECKELLTPEARK-----LEQOVLDCLNKATDE-----ERKKCL 672
Qy      679  SEIPLTHSVSELVRLCLRRSDVOESEGSDTDNNKDSAAFEDNEVQDFLEKLETSBFF 738
Db      673  KDLPKDL-----QSDILAK-----ESLKAYKDCVSOAKTEAEKKECKLLTPEAK 717
Qy      739  ELTSEKLOILITL-----CHRIILTVSYVDHMETRQMSAELWELKRLAV 783
Db      718  KLLBEAEKESVKAYLDCVSOAKTEAEKKECKLLTPEAKKLEAEKSVKAYL--DCVSR 775
Qy      784  LKEENDKKRAEK-----QKKEKEA-----KN-----KENGKVENGLGTRDKRIVKPEP 829
Db      776  ARNEKEKECKELLTPEAKKLEQALDCLKNAKTDEKCKLCKDLKDLKQKVLAK--- 832
Qy      830  QVDTAEADMI SAVKSRRLALAIQAKERFIQEREMKVKLERQAEERIRKHAAAEKAFQE 889
Db      833  -----ESVKAYLDCVSOAKTEAEKKECKELLTPEARKLLEAEK-----SVKAYLD 878
Qy      890  GIAKAKLVMTPTPTGTDNRHNRVWLFSDVPLFIEKGMVHDSIDYRFNHHCQHTVSGD 949
Db      879  CVSOAKTEAEKKECK-----LLTPEARKLL----- 904
Qy      950  EDYCPRSKKNLGNKWSMNTQHTGATEVAVETTTPKQGNLWFLCDSQKELDE----- 1002
Db      905  EXAKESVKAYL-----DCVSOAKNEAEKKECKLLTL-ESKKLEAEKSVKA 952
Qy      1003  LNLCLHPQGIRESQKLERKRYQDIHSHIHLARPNLGLKSCD-----NQELL 1052
Db      953  YLDCV-----SOAKTEAEK-----KECEKLLTPEAKKLEQOAL 986
Qy      1053  NPLRSDLIEV-ATRLQKGLGVVETSEPEARVISLEKLDKDFGECVIALQASVIEKFLQ 1111
Db      987  DCLKNAKTEADKKRCVK-----DLPLQDKVLAKESLKAYKDCVSKARNEKEKKECK 1040
Qy      1112  FNAPKQKERRKQSDS-----AKTEE-----VDEEKKMVVEAK----- 1144
Db      1041  LUTPEAKKLEAEKSVKAYLDCVSOAKTEAEKKECKELLTPEARKLLEAEKSVKAYKD 1100
Qy      1145  -VASAL-EKWKATATREAOQTFSRMHVLLGMLDACIK- WDMSAENARCK- KVCPPKGGEDDKLI 1200
Db      1101  CVSKARNEKEKKECKELLTPEAKKLEQOVLDCLNKAKTEADKKRCVKDLPK--DLQKV 1158

```


[illegible]

Db 2993 SP--YENVPQSFFSSEBSKTDNDANHTTSHFSSEYVSVITITSPVEDVWVASSSGTVLS 2950

Qy 1442 KHLPGHPVRRKRKKPPDLAEDGDS-----EP-----EAVGQ-----SRDEDRR 1482

Db 2951 KESNPEGQDIKESQLESTLWENQSDSVSSSEPTMSATTTVVGEQISKVLIITKDVSDD 3010

Qy 1483 SREABIOEWLQDTSYSAKINSKDHNCFMMLVNTQFCMALDIT 1525

Db 3011 S-WSEIRE--DDEAFEARVKEEQKIFGLMVDVRSQSGITPDT 3049

RESULT 32

A45973

trichohyalin - human

C:Species: Homo sapiens (man)

C:Date: 03-May-1994 #sequence_revision 01-Mar-1996 #text_change 09-Jul-2004

C:Accession: A45973

F:Lee, S.C.; Kim, I.G.; Marekov, L.N.; O'Keefe, E.J.; Parry, D.A.D.; Steinert, P.M. J. Biol. Chem. 268, 12164-12176, 1993

A:Title: The structure of human trichohyalin. Potential multiple roles as a functional E-ed (cross-linking) protein.

A:Reference number: A45973; MUID:93280194; PMID:7685034

A:Accession: A45973

A:Molecule type: DNA

A:Residues: 1-1898 <LEE>

A:CROSS-references: UNIPROT:Q07283; GB:109190; NID:G292835; PIDN:AAA65582.1; PID:G292836

A:Note: authors translated the codon AGG for residue 1714 as Pro

C:Comment: Trichohyalin is a protein of the medulla of the hair and of the inner root sheath. Covalent modifications to this protein include conversion of arginine to citrulline and C:Genetics:

A:Gene: GDB:THH

A:CROSS-references: GDB:136223; OMIM:190370

A:Map position: 1q21-1q21

C:Superfamily: trichohyalin; calmodulin repeat homology

C:Keywords: calcium binding; citrulline; EF hand; hair; tandem repeat

F:49-81/Domain: calmodulin repeat homology <EF2>

Query Match 3.4%; Score 274; DB 1; Length 1898;

Best Local Similarity 18.4%; Pred. No. 0.00014;

Matches 192; Conservative 192; Mismatches 451; Indels 206; Gaps 30;

Qy 483 KENDREDKRSALSVCISKTARLLSGEDRARLPBELRLSVOKRYELLEHKRWASMSSEQ 542

Db 288 RERQEEQQQRLR--REQQLRKKEERREQQERREQQERREQQERREQQERREQQERRE 345

Qy 543 RKE-YLKRRRELKXK---LKEKAKERRKEML-----ERLEKQRYEDQELTKNLPAP 593

Db 346 RREQQLRREQEERREQQRLRREQEERREQQRLRREQEERREQQRLRREQQLRRR 405

Qy 594 RLVDTPPEGLPNTLFGDVAMVVFVFLSCYSGLLLPDQVPTAVSLMEALSADKGGFLYLNR 653

Db 406 QQLRREQQLRR-----QQLRREQQLRRREQ-----QQLRREQQLRRREQ-----QQLRR 434

Qy 654 VLVILLQTLLODEIAEDYGEKMGKLSIPTLTHSVSELVRLCLRRSDVQSESGSDTDDN 713

Db 435 EQQLRREQQLRREQEERREHQHEQERREQLRKEQEERDMLKR---EBETERHEQEERR 491

Qy 714 KDSAAFENEDVDEFLKLETSEFFELTSEEKLIITLALCHRLMTYSVQDHMETR-QQM 772

Db 492 KQQLKRDQEEERERWLKLE-----BEER-----REQQRREQQL 526

Qy 773 SAEWLKERLAVLKEENDKRAEKQKRWKEMAKNKKVNGKGLKTDK-----KRIVKF 827

Db 527 RREQEERREQLRKQEERREQLRRLSEQQQLRREQEERLEQLLKREKEKLEQERREQRL 586

Qy 828 EPQVDTEADMTISAVKSRLLAIAQAKKREIQREMKVKLERQABE-----ERIKKHA 883

Db 587 KREQEERRDLKREERREQQRLRKQEERLEQLRKEEVERLEQEERREQLRKEEPEE 646

Qy 884 EKAFQEGTAKALVWRPTIGTDNRNRYLWFSDEVPGLFIEKGWVHDSIDYRFNHCKD 943

Db 647 ERRELLKSEQEERREHQERREQQERREQLRKEE-----EERLEQLRKREHEERRE 701

QY	1042	LKSCDQNELNPLRSLDLIEVATRLQ-----KGLGYVEETSEBEARVISLEKLD	1092	QY	133	ISNVPADS-LIRTERPPNKEIVRYFIRHNALRAGTGENAPWVVED--ELVKYSLPSKFS	289
DB	933	-----LNSMQQLIOAQNTLQKNEKEBQQLQGNINELQSSQKKK--QIEALQ-	979	DB	1261	QAMPLADSQAVREQURQOELEIEIRH-----GEK-----VEBQRFQAYINAIQDY	1309
QY	1093	FGECVIA-LQASVKKFL-----QGPMAKQKRKLQS--EDSAKT-----	1130	QY	290	DFLDLPYKMTLPNSTRTKNTGSPDRKPSKSKTDNSLSPLNPKLWCHVHLKK--SLS	347
DB	980	-GELIAVLQKTELENKLOOQLOTOAAQELAAEKEKISVLQNNYKSOETFKQLQSDFYGR	1038	DB	1310	ELQLVITYKAQL-----EPVASPAKPKVQSGS-ESVIOEYVLDLRTHYSELTLT	1357
QY	1131	-----EEYDEEKKWVEEAKVASALEKWKTAIREAQT	1161	QY	348	GSPK-KVNSKNSKSPPEHLEMMKMMSPNK-----LHTNFHIPKKGPPAKPKGKHSD	399
DB	1039	ESELLATRODLKSVEKLSLAQEDLISNRNQINQNKLIQELTKATAKTLQDSAKKEQQL	1098	DB	1358	SOYKIFSETLRRMEERLAEQQAERERLAEVEAALERQOROLAEAAKAQAAREA	1417
QY	1162	FSRMHVLLGMLDACLKWDMSAENARKV-----CPKKGDDKLILDCBCKNAFLPCL	1214	QY	400	KPLAKAGSKGLNGQKSTGNSKSPKGLKTPKTKMKOMTLLDM-AKGTQKMTAPRNSG	458
DB	1099	QBRKALODIQEKSILKELYNKSLAEIBEIKRQKEITKL-----NEELSKHLES	1155	DB	1418	KELQORMQEE-VVRRERAAVDAQQKRQSTQBELQOLROSSAEIQKA-----ROAE	1468
QY	1215	RPALVEVPDG-----EWQCPA-COPATARRNSGRNYTEESASEDESDEBEE	1264	QY	459	CTPTSTSPKHPLPPAALHLIAYYKENKDREDKGSALSCVISKTAARLSSBDRALPPEL	518
DB	1156	K-EITNLKDAKQLLQOKLEQKADSLKAAVEQEKRNQOILKDQVKKEEELKKEFIEK	1214	DB	1469	AAERSRLAIBEEIRVVRLQLEATERQRGGAEGELQAL-----RAR-AEEA	1512
QY	1265	EEEEEEEDYEAGLRLRPRTIRGHSVIPPAARSGRRPGKPHSTRSQPKAPPVDD	1324	QY	519	RSLVOKRYELLEHKRQWASMSBQKEYLKKRBEKKLKKLXKAKERREK-----EMLERL	574
DB	1215	EAKLHSEIKEV-GMKHEENEAKLTMOITALNENLGTVKWQSSQRRVSELEKQTD	1273	DB	1513	EAQKQAOEBAERLRQV-QDESQRK---RQAEVELASRVKAETEAAREKORALQALBEL	1568
QY	1325	AE-----VDELVLQTKSSRSQSLQKC-----EELHKIVKYR-----	1359	QY	575	EKQKRYEDELTKNLPAPRLVDTPEGLPNTLFGDVAMVVEFLSCYSGLLLPDAQYPTA	634
DB	1274	LRGEITAVLATVQNNQDERALLL--RCLKGEIEKLTQTKLELQKLDNTTAVQELGR	1331	DB	1569	RLQAEABERRLRQAEVERARQV-----QVALETAQSAEAELOSKRASFAKBT	1616
QY	1360	---FSWPPFPVTRDEADYDVITHPMDFTQVONKSCGSGSYRSV	1401	QY	635	VSLMEALSADGGFLYLRNLVILQTLLODEIAEDYGELOGMKLSIEPLTLHVSSELVRL	694
DB	1332	ENQSLQIKHTQALNKWAED-----NEVQNCWACGKGSFV	1366	DB	1617	AQLERSLQEEHVAVAEAREEARRAQQAEEAREEARELERWQL-----KANEALRL	1671
RESULT 34				QY	695	CIRSDVOESESGETDDNK-----DSAAFEDNEVDQEFLEKLETSEFFE	739
G02520				DB	1672	RLQAEVQAQKSLAQAEAEKQKEAEREARRRKAEOAVRQRELAEOLEK--ORQLAE	1729
Plectin - human				QY	740	LTSEBKQLTALCHRLIMTYSVQDHMETRO-QMSAELWKERLAVLKEENDKKAERKOKR	798
C:Species: Homo sapiens (man)				DB	1730	GTAQORLAAEQEL-----IRURAEQEOQRQLEBELARLQRE---AAATQKR	1777
C:Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 20-Aug-1999				QY	799	KEMEAKNKENGKVENGLKTKDKRKIRVKFEPQVDTAEADMSIAVSKSRLLAIQAKKEREI	858
C:Accession: G02520				DB	1778	QELEA---ELAKV-----RAEMEVLLASKARAEESRSTSEKSKQRLAEACGRFREL	1826
R:McLean, W.H.I.; Smith, F.J.D.				QY	859	QEREMKV-----KLERQAEERIRKHAARAKAFQEGIA-----KAKLWMRT	901
submitted to the EMBL Data Library, March 1996				DB	1827	AEEARLALAEAKRQRLAEEDAAARQRAEAEVLAEKLAIGEATRLKTEAEIALKEK	1886
A:Reference number: H01385				QY	902	PIGTDNRNRYWLPFSDV-----PGLFTEKGWVHDS	932
A:Accession: G02520				DB	1897	EAENELRR---LAEDFAFORRLLEEQAQAHKADIEERLAQLRKASDSELERQKGLVEDT	1943
A:Molecule type: mRNA				QY	933	IDYFRNHKCDHTVSGDEDYCPRSKKNALGNKSNMNTQHGTAATEVAVETTTTQKQNLWF	992
A:Residues: 1-4574 <MCL>				DB	1944	L--RQRQVEBEILALKASF---EKAAGK-AELELELGRIRSNADTLRSKE-----	1990
A:Cross-references: EMBL:U53204; NID:g1477645; PIDN:AAB05427.1; PID:g1477646				QY	993	LCDSQKELDELINCLHPQGIRESQKLEKRYQDIHSHILARKPNLGLKSCDQNOELL	1052
C:Superfamily: plectin; alpha-actinin actin-binding domain homology; ribosomal protein s				DB	1991	---QAE-----EAQRQLAAEERERRREAEERVO-----KSLAAEEE---	2026
F:68-283/Domain: alpha-actinin actin-binding domain homology <ACT>				QY	1053	NFLRSDLIEVATRLQKGGYVEETSEFEARVISLEKLDGECVIALQASVKKFLOGF	1112
Query Match 3.4%; Score 267; DB 2; Length 4574;				DB	2027	-----AARQKAAL---EEVERLKANVTEARLRERAEQESARQ-----LQ---	2064
Best Local Similarity 18.6%; Pred. No. 0.00077;				QY	1113	MAPKOKRKLQSDSA-----KTEEVDEEKKWVEEAKVASAL-----EKWKTAIREAQTF	1163
Matches 303; Conservative 238; Mismatches 612; Indels 458; Gaps 66;				DB	2065	LAQEAQKRLQAEKKAHAFVQKQEQELQOTLQEQSVLDQLRGAEAAARAAAEAE---	2121
QY 20 EPFFTIPHTQEAFT-REYE---ARLER-----YSERIWTCK-----STGSSQL--T				QY	1164	RMHVLLGMLDACIKWMSAENARKCVCKPKGDEDDKLLILDCBCKNAFLHFLCLRPALYVDP	1223
DB 1075 EKVLALPEPSPAPTLRSELTGLKLEQVRSLSAIYLEKLTISIVIRGTQGAEEVLRA				DB	2122	-----EARVQAREAAQARRQV-----EEAER-----LKQSAEQAO	2153
QY 62 HKEAMEEEOEVALLEKEEPFAPWVEKLVLEWVHNHTASLEKLVDTTAWLEIMTKYAVGEC				QY	1224	GEWQCPA-----COPATARRNSGRNYTEESASEDESDEBEEEBEEDY	1275
DB 1135 HEEQLKEAQAVPATLPE-----LEATKASLKXRAQAEAAQOPTFDALRDE---							
QY 122 FEVGKEMLVKIVKTHPLEKVEDEATEKKSQDCAKSPSSDKENS---SOIAQDHQKKT							
DB 1180 -----LRGAEVGERLQQRHGE-----RDLEVERWRERVAQVLERWQA							
QY 179 VVKEDEGRRESINDRARR-----SPRKLPTSLKKGKRWAPPKFLPHKYDVKLQNEDK							
DB 1218 VLAQTDLRLEQLGRQRLRYRESADPLGAWLQDARRR-----QEQI							

Db 2154 ARAQAQAAAEKLRKEAEQAEARQAQAEQAAALRQKQADAEKHKFAEQTLRQKAQVEQ 2213
QY 1276 EVAGLRPRKPTIRGKHSVIPPAAARRRRPGKKPHSTRSQPKAPPVDDAEVDELVLQTK 1335
Db 2214 ELTTLRLQLEETHQKML-----DEQLRLKAEAT 2244
QY 1336 RSSRROSLQKCEILHKI-----VKYRFSWPPFPFVTRDEADYDVIHPDMDF 1386
Db 2245 EAARQRS-----QVEEELFSVRQMEELSCLKARIEAENRALILRDXD----- 2287
QY 1387 QTQVQKSCSYRSVQEFITDMKQVFTNA-----EYVNCR-----GSHVLSMW 1430
Db 2288 -----NTQRFLOEAEKMQVAEEAARLSVAQAQEARLQRLAEEDLAQORALAEKM 2338
QY 1431 KTEQCLVLLHKLPGHPYVRRKPKFPD-----RLAEDEGDSPEAVGQSRDSD-----RR 1482
Db 2339 LKEMQAVQATRLKAEALLOQKELAQEARLQEDK-----EQMAQQLAEETQGFQR 2393
QY 1483 SREAEIQEWLQ 1493
Db 2394 TLEAERQRLQLE 2404

RESULT 35
B85431
trichohyalin like protein [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 09-Jul-2004
C:Accession: B85431
R:anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Spring
Nature 402, 769-777, 1999
A:Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.
A:Reference number: A85001; MUID:20083488; PMID:10617198
A:Accession: B85431
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1432 <STO>
A:Cross-references: UNIPROT:O23230; GB:NC_001268; NID:g7270600; PIDN:CAB80318.1; GSPDB:C
C:Genetics:
A:Gene: At4g36520
A:Map position: 4

Query Match 3.3%; Score 265.5; DB 2; Length 1432;
Best Local Similarity 20.3%; Pred. No. 0.00023;
Matches 293; Conservative 206; Mismatches 506; Indels 439; Gaps 70;

QY 61 THEAEWEQAEVALLKEBFPWYKELVLEMHHTASLEKLVDTAWLIMTKYAVGERC 120
Db 268 THPTGIPPPSRAAPILKSDFRS-----SASNSKTTGSGQVDS--SSPTFFDVEVDA 318
QY 121 DFEVGEKMLKVIKVIHLEKVEDEATEKSDGA--CDSPSDKXNSQIAQDHQKKT 178
Db 319 NSAAVREMLKAE--AKL-----KSAKLLERKRDVARTSTPSSTK--NRRIEKGSSYA 370
QY 179 ----VVKED--EGRESIND--RARRSRKPLTSLKGERKWAPKFLPHKYDVKLQVEDKI 232
Db 371 AGLPVVKSINIDVARKSLKRGSKSLSSQGSATSDGNDW-----KEANNQVY 418
QY 233 ISNVPADSLIRTPPNKE-----IVRYFIRHNLRAAGTGENAPWV----- 274
Db 419 -----ELVRTLPRNADENSGKDVSIPLNAEFFDQELTWAG---NVDWEKQRRRAK 468
QY 275 ---EDELVKVSLPSKSFDFLLDPKYWTNLPSTYKNTGSPDRKPSKSTONSLS--S 330
Db 469 DREDHEARK--LP-----KHSQT--RKLSSRRKHENKLAEKA 502
QY 331 PLNPKL--WCHVHLKSLSSGSLPKVKNKSNKSPREEHLEMMKMSPNKLHTNFIHPKKG 388
Db 503 PREPKLEKSRHVMGNDLPDHGGIVKH--RNLLKPEE-----NKLFT----- 542
QY 389 PPAKPGKSHDKPLKAKGRSKGLNQKSTGNSKSPKGLKTPKTKQKQMTLLDMAKGTQ 448

Db 543 ---EKPAKQK--KELLCEETKRIQNOOL---DKTKHQKAAET-----NOECVYDVEQNAR 590
QY 449 KMTAPRNSGGTPTSSKPHKLPAPAAHLIAYYKENDREKRSALSALCVISKTARLSS 508
Db 591 KUREALGNB-----STLEVSVELNGNGKMMEMRSOSETKLNPLKRMEE 634
QY 509 EDR---ARLPEELRSIVQKRYELLEHHKRW--ASMSEORKEYLKKKREBLKK-----KUK 559
Db 635 ETRIKEARLREENDRRERVAEKAENEKRLKAALQEKEERKIKBEAREKAENERRAVEAR 694
QY 560 EKA--KERRKEKMLE--RLEKQKRYEDELTKGNLPAPRLVDTPEGLPNTLFGDVAMVVEFL 617
Db 695 EKAEQERKQEQLELQKAEFEKEEENRRMRREAFALQEKE----- 738
QY 618 SCVSGILLPDAQYPIITAVSLMEALSADKGGFLYLRNLVILLOTLQDAIDAEYGLGK 677
Db 739 -----RIKEAREKEENE--RRI-----KEAREKALEQR 765
QY 678 LSEIPLTLHSVSELVRLCLRRSDVQESGSDTDNDKDSAAAFEDNEVDQEFLEKLETSF 737
Db 766 LK---ATLE-----QEEKE---RQIKERQEREENERRAK--EVLQAE-- 800
QY 738 FELTSEELQILTALCHRIILMTYSVQDHHMETQQMSAEKWLKERLAVLKBENDKRAEKQK 797
Db 801 ----NERKLK-----EALQ-----KENERRLKETREKENKKL 831
QY 798 RKEAKNKENKENVGLKTKDRKKRVKFEQVDTAEADMISAVKSRLLAIQAKKERE 857
Db 832 REAILEEKEKLLIE--AFERAIEIRRL-----KEDLEQEM-----RMLQEAERE 877
QY 858 IQEREMKVLQRAEERIRKHAAAKAFQGIKAKLVMRRTPITGTDNRNHNRYLWFS 917
Db 878 RLHRE-----NQHQENERKQHEYSGESEDEK--ERDCEMEKTCETTKEAHEQ---SSN 928
QY 918 EYVGLFIEKGWVHDSI--DYRFNHHCKO-----HTVSGDEDYCP----- 954
Db 929 ESLSDTLEE---NESIDNDVSVNKKQKEBEGTRQRESMAETCPWKVFEKTLKDAQSQEG 985
QY 955 -----RSKANILGNASMTQHGATATAVAVETTTPKQONLWFLCDSQK--- 998
Db 986 TNEMDADTRLFERNEETPRLGNGCNGQNGSGEESTSVTENIIGGKL-----EOKSKN 1040
QY 999 -ELDELLNCLHPQGIRESOLKREKRYODIHTSHLARKPNLGLKSCDGNOLLNPLRS 1057
Db 1041 SETSKDASVLK---RVSGLKTVEVERLEDVVGVRDQDNPE---ESKSAPTYSYCF--RN 1092
QY 1058 DLIEVATRLQKGLGYVETSEFEARV-----ISLEKLKDFGECVIALQASVTKKFLQ 1110
Db 1093 HEYKFTHQOQERNIVETQAGLNQDAKVERPLPSRVSVQREKE----- 1134
QY 1111 GPWAPKQKR--RKLQSEDSAKTEVDEBKQWVEAKVASALEKWKTAIREAQTFSRMHVLL 1169
Db 1135 ---AERLKRERDLEQLRKVEEREREREKDRMA-----FDOR----- 1172
QY 1170 GMLDACIKWDSMAENARCKVCPKGEDDKLILDCENKAFHLFCLRPALYEPDGEWQCP 1229
Db 1173 ALADARERLEKACAREKSLP-----DKLSM-----EARLR 1204
QY 1230 ACQATARNRSGRNYTEESASEDSEDESEDESEDESEDESEDESEDESEDESEDESEDE 1289
Db 1205 AERAVERATSEARDRAAEKAA-----FEARERSVSDKQSQSQSGFFGERMEISLSDK 1259
QY 1290 G-KHSHVIPPAAAR-----SGRRPKKPHSTRSQPKAPPVDDAEVDELVLQTKSSRRQ 1341
Db 1260 QFQNSVSGASRYQDSHGTEGSPQYTSRLERHQTADRVAIAET--LDTEVGRWSSGK 1317
QY 1342 SLELQKCEILHVIKRYFSW---PFREPVTRDEADYDVIIT---HPMDFQ---TVQN 1391
Db 1318 EGNIRALLSTLOYILGPESGMOPLPTEVITSAAVRAYRKATLCVHPDKLQOORGANIHQ 1377
QY 1392 KCSC 1395
Db 1378 KYIC 1381

RESULT 36

T20488
 hypothetical protein F01G4.1 - *Caenorhabditis elegans*
 C:Species: *Caenorhabditis elegans*
 C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
 C:Accession: T20488; T24527
 R:Harris, B.
 submitted to the EMBL Data Library, January 1996
 A:Reference number: Z19281
 A:Accession: T20488
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1474 <W1>
 A:Cross-references: UNIPROT:Q19106; EMBL:Z68341; PIDN:CAA92768.1; GSPDB:GN00022; CESP:F01G4.1
 A:Experimental source: clone F01G4
 R:Kershaw, J.
 submitted to the EMBL Data Library, January 1996
 A:Reference number: Z19903
 A:Accession: T24527
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1474 <W12>
 A:Cross-references: EMBL:Z68751; PIDN:CAA92978.1; GSPDB:GN00022; CESP:F01G4.1
 A:Experimental source: clone T05E11
 C:Genetics:
 A:Gene: CESP:F01G4.1
 A:Map position: 4
 A:Introns: 152/3; 227/3; 303/2; 564/3; 567/3; 1117/3; 1286/3; 1448/3
 C:Superfamily: human SNF2alpha protein; bromodomain homology
 F:1208-1263/Domain: bromodomain homology <BRO>

Query Match 3.3%; Score 265.5; DB 2; Length 1474;
 Best Local Similarity 18.3%; Pred. No. 0.00024;
 Matches 294; Conservative 231; Mismatches 549; Indels 529; Gaps 68;

QY 139 PLEKVDDEATEKSDGACSPSDKENSQIAQDHQKKTETVVEDEGRRESINDRARRSP 198
 DB 14 PPQAGEVVAQAGDGA-----QQQELTISKLENAITSMEEQGLQN- 54
 QY 199 RXLPTSLLKGERKWAAPPKFLPHKYDVKLQNEKIIISNVPADSLIRTPPNKEIVRYFIR 258
 DB 55 -----DIRHAKAVLLKQKLSGLP-DAVPGQENGNOIIP--AQ 91
 QY 259 HNALRA-----GTGENAPW-VVEDELVKYKSLPSKFSDFLLDPYKYMTLNPSTYGRKN 309
 DB 92 LNQLRAQVSAVRLARNEQVPANLIADAVM---LRPKVTLLPEPYEY-----PGEAENG 143
 QY 310 TGSP-----DRKPSKSKTNSLSPLNP--KLWCHVHLKSLSGSPKLVKNSKNSK 360
 DB 144 EKLPLYDLMKIFNLHQIRCNRPRTTISVPSGIDVPVGLKQRENMIQNRIGLRMKLLANLIPAD 203
 QY 361 SPEEHLEENMKWMSPNKLTNPHIPKGGPPAKPGKHSDKPLKAKGRSK--GILANGQKST 418
 DB 204 IFDH-----MKKAEIELRALRV-----NLQTVRSEVMACLRDITL 242
 QY 419 GNSKSPKGLKTPKTKMKQMTLLDMAGKTQKMTAPRNSGGTPTSSKPHKLPPAALHL 478
 DB 243 ETALNPYAVRRTKRQSLREARVTEKLEKQKMEQ-----ERKRQKHDTLMQA---I 291
 QY 479 IAYYKNDREKRSALSCVISKARLL--SSEDRARLPPELRS-----LVQRYE-- 527
 DB 292 IQHGEFKEYH--RNLLKMAKSRKAVMTYHQNREKKEDEIRNEKLRMQKLMQDEEGY 350
 QY 528 --LLEHKRWASMSRQKEYLKKREELKKLKEKAKERREKEMLERLEKOKRYEDQEL 585
 DB 351 RALLDEK-----DQRLVYLQQDETVDSLSLVROH-----QNTKKKKCKEDKKI 397
 QY 586 TGKNLPAPFLVDTPEGLPNTLFGDVAMVVEFLSCYSGLLLPDAQYPIITAVSLMEALSADK 645
 DB 398 EKGN-----QWDEEAVHVRERSTGKALTGD- 423

QY 646 GGFLYNLRVVLVILLQTLQDEIAEDYGELGMKLSSEIPLTLHSVSELVRLCLRRSDQVES 705
 DB 424 -----QAPKTEEI-----EFWLETHPEYEVIPRDQLSDDEBEE 457
 QY 706 EGS--DTDNKDS--AAPEDNEVQDEFLEKL--ETSEFFELTSEBKQLTALCHRIILMT 759
 DB 458 EEAPEPEEKDDQVAGWDEETKAKMILEKARNEDEYDQTKKQ----- 502
 QY 760 YSVQDHMETROQMSAELWKERLAVLKEENDKKRAEKQKRKEVE-----AKUKENGKV--E 812
 DB 503 --MADYYATAHKIKIKEKVVVKQH--TTMGGGDPNLLLKPYQIKGLEMMVSYNNNLNLADE 559
 QY 813 NGLGKTDRKKLVKPEPOVTEAEDMISAVKSRRLLAIOAKKERIOQREMKV----- 865
 DB 560 MGLGKTIITISLVLYLMEVKQNNPGYIVIP-----LSTLSWQNEFAKWAFSVTTIIVKG 615
 QY 866 -KLERQAESEERIRKHAAAEKAFQEGIAKAKLVMRRTPIGTRNHNRYWLFSDVPGFLFI 924
 DB 616 TKDARRRVEGQIRKGAFNVLMTTYEYVIKELALGKI-----RWKY-----MLI 659
 QY 925 EKGWVHDSIDYRFNHHCKDHTVSGDEDYCPRSKZKANLGKNSAMNTQHGTAATEVAVETTP 984
 DB 660 DEG-----HRLKNHN-----C-----KLTLMLNGFFFAQHR-----LLLTGTP 692
 QY 985 KQGG--NLWFLCD-----RSQELDELINCLH----- 1008
 DB 693 LQNLKPELWALLNLLPFISSCGTFEQWFNAPFATGGEKVELNKEETMLIRRLHKLVR 752
 QY 1009 PQGIR-----ESOLKERLE-----KRYQDIIHSIHLARKPNLGLKSCDGNQELLN-- 1053
 DB 753 PFLRLRLKKEVESQLPDKTEYVIKCDQSALQKVIYR-HMQGLLLDAXSSGARSALMNTV 811
 QY 1054 -----FL-----RSDLIEVATRLQKGLGVVEETSEFEARVI 1085
 DB 812 VHLRKLCHNPLFPNIEDSCRAYWKVNEVNGTDLMRVAGKL-----ELLDRI- 858
 QY 1086 SLEKLKDFGECVIAL-----QASVIKKFLQ-----GFMAPK- 1116
 DB 859 -LPKUKATGHRIIMPFTQMSMNIFEDFLNFRYYTLRLDGSTKPDGERGDLTLTQFNAPNS 917
 QY 1117 -----QKRKLQSEDSAKTEEVEDEKMWVEBAKV 1145
 DB 918 DLFLMLSTRAGGLGLNLQATDVTIIPDSWNPHQDMQAQDRAHRIQKKEVRLRLITA 977
 QY 1146 ASALEKWKTAIREAQTFSRMHVLLQMDACIKWMSAENARCK-----VCPKKG- 1195
 DB 978 NSVEEKILAAARYKLNVDKVIQAG-----KFDQSTGAERKQMLEQIIQADGSEBEE 1031
 QY 1196 --DDKLI--LCDECNKAFHLF-----CLRPAL--YVPPGEMWCPACQ 1232
 DB 1032 EVPDDETVQWVARSEEFNIFQSMIDIRREANQLHRKPLLEEHIIPDDILKLSFDY 1091
 QY 1233 PATARNRGRNYTEESASD-----SEDDSEDEEEEEEDEEYEVAGRLRP 1284
 DB 1092 EEMERAREEGREVVDQTNQRRRRREVYSSDLSLDEQFMQVEVEEDNNQVAERKKQ 1151
 QY 1285 RKTIRGHSVIPPAAARSGRPKGPHSTRSQPKAPPVD--DAEVDLVLQTKRSRRQS 1342
 DB 1152 RKR-----KMAGLDENDSDMDVYLQHKKKTDPE 1181
 QY 1343 LELOKCEILHKIKVYR-----FSWPREPVTRDEADYDVITHPMDFTQVQNKCSG 1396
 DB 1182 L-AEKINEMLOVILEYKNEDEGLIADVFTQTPTRKELPDYQVLSKPMDFDRINKIETG 1240
 QY 1397 SYRSVQEFELTDWKQVFTNAEVNCRGSHVLSQWVTEQCLVVLHKL-----PGHPYV 1450
 DB 1241 RYTWMEELNDMDNLLVNAQIYNEEGSEI---YVSSE--TIGKLWKEQYDFMFPKPVE 1295
 QY 1451 RRRKXKFPDRLAEDGSDPEAVGQSRDED--RRSREAIQEWL 1492
 DB 1296 BPVKKKEPS--TPSTSSSRPSTSGTPSVSDLQRTQQAQOML 1336

RESULT 37

A59282
nonmuscle myosin II heavy chain A - African clawed frog
C:Species: Xenopus laevis (African clawed frog)
C>Date: 09-Jun-2000 #sequence_revision 09-Jun-2000 #text_change 09-Jul-2004
C:Accession: A59282
R:Bhati.Dey, N.; Taira, M.; Conti, M.A.; Nooruddin, H.; Adelstein, R.S.
Mech. Dev. 78, 33-36, 1998
A:Title: Differential expression of non-muscle myosin heavy chain genes during Xenopus e
A:Reference number: A59282; MUID:99077683; PMID:9858676
A:Accession: A59282
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-1964 <BHA>
A:Cross-references: UNIPROT:O93522; GB:AF055895; NID:g3660671; PIDN:AAC83556.1; PID:g366
A:Experimental source: Cell line XTC
C:Superfamily: myosin heavy chain; myosin motor domain homology
F:84-764/Domain: myosin motor domain homology <MO>

Query Match 3.3%; Score 265.5; DB 2; Length 1964;
Best Local Similarity 18.9%; Pred. No. 0.00034;
Matches 266; Conservative 206; Mismatches 463; Indels 473; Gaps 57;

Qy 29 QEAFTREEEARLERYSERIWTCKSTGSSQLTHKEAWEEQVALLK--EEFPAYEK 86
Db 911 QLEEEILHDLSEVBEER-----CQILQTEKKKQQNVILEEQLEEEEAARQK 961

Qy 87 LVLEMVHNHTASLEKLVDTAWLBITKAYAGEBCDPEVGEKMKLVKVIKIHLEKVDSE 146
Db 962 LQLEKTAB-AKMKLEEDV-----LVLEQNTLKEKKLYE-----ERIAEF 1004

Qy 147 ATEKSDGACDSPSKENSSQIAQHQKQKTYVVEDEGRRESINDRARRSKPLTSLK 206
Db 1005 TT-----NLTEEEKSKSLAKLNKHETWISD-----LEERLRE-EKORQELE 1047

Qy 207 KGERKWAPPKFLPHYK-----DVKLO--NEDKIISNVPADSLIRT-ERPPNKEIV 253
Db 1048 KTRRKLEGSDTLHDQIAELQAIAELQLAKKEEELQA-----ALARAEEAAQKNLA 1102

Qy 254 RYFIRHNALRAGTGENAPWVVEDELVKYSLSKFSDFLLDPYKMTLPSTKRKNTGSP 313
Db 1103 LKKIRE--LESQIG-----LQBDLESRA-----ARNKAEKQKRDIGE 1139

Qy 314 DRKPSK---KSKTDNSLSPLNPLWCHV-HLKSLSGSPLKVNKSNKSPPEHLEEM 369
Db 1140 ELEALKTELEDTLSDATQQLATKREQEVTHLKKTL-----EDEARTHGEQIQEI 1190

Qy 370 MKMSPNKLHTNPHIPKGPAPKPKGSHSDKPLKAKGRSKGIUNGOKSTGNSKSPKKGK 429
Db 1191 RQ-----KHQO---AVBELSEQLRQTKRLKGNLEKAKQAQLE 1223

Qy 430 TPKTKM--KOMTLDMAGKTQKMTAPRNSGGTPRTSSKPHGLPPAALHLIAYYKENKD 487
Db 1224 GERNEIANEVKTL-----QKGDSHKKRKKVEAQLQELQVKVTGDRVRSE 1270

Qy 488 REDKRSALSCVISTARLLASSEDRLRP-----EELRSLVQKRYELLEHKRKMWSMSBEQ 542
Db 1271 LSEKANRLQVELDNVNSLSQSDSKSLKGDFTLESQFQDAQELLQ----- 1318

Qy 543 RKEYLKKREELKKLKEKAKEREKEMLERLEKQKRYEDQELTGKNNLPAFLVDPPEGL 602
Db 1319 -----EETRQKLSFSTKLQMEDEKNGLEQL-----BEEERAKCNL----- 1355

Qy 603 PNTLFGDVAMVFEFLSCYSGLLLPDAQYPITAVSLMEALSADKGGFLYLNRLVILLQTL 662
Db 1356 -----CKQ-----ISTLQSQ 1365

Qy 663 LQDEIAEDYELGEMKLSEIPLTHSVSELVRLCLRRSDVQSESGSDTDNDKDSAAFEN 722
Db 1366 MTD-----MKKKDENVGSLTEVEEL-----KKKLQKOLEAVNQRFEEKAAAYDK- 1410

Qy 723 EVQDEPLEKLETSFPELSEKQLQILTALCHRLMTYSVQDHMETRQQMSLEWKERLA 782

Db 1411 -----LEKTKTR-----LQQELDDISVDLDHQ-----ROIIVSNLEKQKKKFDQLGEEKA 1455

Qy 783 VLKEENDKRAEKQKQKMEAKNKEN-----GKVENGLGKTDRKKRIIVKFEQVQVTEAE 836
Db 1456 I-----SAKYADERDSEAEAREKETKALALAPALEAL---ZAKAELERLNKQLATME 1507

Qy 837 DMISA-----VKSRRLLAIOAKK-----EREIOBRE-----MKV 865
Db 1508 DLVSSKDDVGKSVHELEKSKRALEQQAEMKMTQLELEDELQATDAKLRLVNLQAMKA 1567

Qy 866 KLER-----QAEER---IRKHKAATAKAFQEGIAKAKLVMRRTPIGTD----- 906
Db 1568 QFERDLQGRDEQSEKKQLVQVQKMEAELEDERKORSALVAARKLEMDLKDLEGQID 1627

Qy 907 -RNHNRYMLFSDEVPGLFTEKGVHCHDHTVRFNHHCKDHTVSGDEYDCPRSKKANLGKNA 965
Db 1628 SSNNKR-----EDAIKOLRKLQAIKD-----VQRELEDTFRASD-DILSQSKES-----EK 1673

Qy 966 SNNTOHGTATEVAVETTPKQGNLWFLCDSQKELDELNCL---HPQIRESQLKERLE 1022
Db 1674 KLKSMBAEMIHMQEELAAAAERGKR-----QAQERDELADEIANSSGKGALALEEKRRLE 1728

Qy 1023 KRYQDIIHSIHLARKPNLGLKSCDGNQELLNPLRSLDIEVATLQKGGLYVVEETSEFEA 1082
Db 1729 SRIAQLEER-----LEEEQGNTELNV-----DLRKKSTLQIDQLNLDLTA 1768

Qy 1083 RVISLEKLDGFCGVIALQASVKKFLQGFMAPKQKRRKLQSDSDAKTEVDEBEKK----- 1138
Db 1769 ERSNAQKNEN-----ARQMDRONKELTKLQEMEGIVKSKFK 1806

Qy 1139 ---MVEEKVASALEKWKTAIRBAQTFSRMHVLLGMLDACIKWMSAENARCKVCPKQGE 1195
Db 1807 ANITALEAKIALEEQLDTETKERQWASK-----QVRRTEKKLK 1845

Qy 1196 DKKLILDCENKAFHLFCLRPALYEVDPGEWQCPACQATARNRSGRNYTESASEDSE 1255
Db 1846 DVLMLVEDE-----RRNS---EYKQQAQKXNV 1871

Qy 1256 DDESDEEEEEE-----EEEDYEVAGLRPRKTIIRGHSHVIPPAAARG 1302
Db 1872 MKQLKROVEAEAEAEQANAMRRKLQRELEDTATETAEIWNREVNTLKK-----LRRGG 1925

Qy 1303 RRPKPKPHSTRSQKAPVDDAEVDEL 1330
Db 1926 GGGGDIITFVTRVRGRKG--VDEASDEDL 1951

RESULT 38
A59252
myosin heavy chain, nonmuscle, form IIB - human
N:Alternate names: myosin type 10; NMHC-B
N:Contains: myosin ATPase (EC 3.6.4.1)
C:Species: Homo sapiens (man)
C>Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 09-Jul-2004
C:Accession: A59252; B61231; G02055
R:Phillips, C.L.; Yamakawa, K.; Adelstein, R.S.
J. Muscle Res. Cell. Motil. 16, 379-389, 1995
A:Title: Cloning of the cDNA encoding human nonmuscle myosin heavy chain-B and analysis
A:Reference number: A59252; MUID:96025307; PMID:7499478
A:Accession: A59252
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-1976 <SIM>
A:Cross-references: UNIPROT:P35580; GB:M69181; NID:G641957; PIDN:AAA99177.1; PID:G64195
A:Experimental source: clone lib Lambda Zap II adult human T-cell library; cell line Ju
A:Note: Between nucleotides 1942-1943 in mRNA encoding human brain MHC-B there is an al
R:Simons, M.; Wang, M.; McBride, O.W.; Kawamoto, S.; Yamakawa, K.; Gdula, D.; Adelstein
Circ. Res. 69, 530-539, 1991
A:Title: Human nonmuscle myosin heavy chains are encoded by two genes located on differ
A:Reference number: A61231; MUID:91316803; PMID:1860190
A:Accession: B61231
A:Molecule type: mRNA
A:Residues: 63-237, 'K', 239-664, 'L', 666-722 <S12>

A:Cross-references: GB:M69181; NID:G641957
 R:Weir, L.
 Submitted to the EMBL Data Library, August 1995
 A:Reference number: H00753
 A:Accession: G02055
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-81 <HEI>
 A:Cross-references: EMBL:U34304; NID:g1143217; PIDN:AAA84880.1; PID:g1143218
 C:Genetics:
 A:Gene: GDB:MYH10
 A:Cross-references: GDB:127350; GDB:G00-127-350; OMIM:160776
 A:Map position: 17p13-17p13
 C:Superfamily: myosin heavy chain; myosin motor domain homology
 C:Keywords: actin binding; ATP; coiled coil; hydrophobic; methylated amino acid; nucleotide
 F:88-771/Domain: myosin motor domain homology <MMO>
 F:178-185/Region: nucleotide-binding motif A (P-loop)
 F:559-572/Region: actin binding #status predicted
 F:633-647/Region: actin binding #status predicted
 F:129/Modified site: N6,N6-trimethyllysine (Lys) #status predicted
 F:184/Binding site: ATP (Lys) #status predicted
 F:701,711/Active site: Cys #status predicted

Query Match 3.3%; Score 265; DB 2; Length 1976;
 Best Local Similarity 18.9%; Pred. No. 0.00036;
 Matches 335; Conservative 263; Mismatches 566; Indels 606; Gaps 81;

QY	15	PUPGEEPTTIPHTQAFR-----TREYEARL-----	RYSERIWTCKS 54
DB	316	PIPGQDKDNFQETWEAMHMGFSHEILSMKLVSSVLQFGNISPKERNTDQASMPEN 375	
QY	55	TCSSQLTH-----KEAWEEQ-----EVAELLKEEPPAWYEK 86	
DB	376	TVAQKCHLLGNVWEFTRAILTPRIKVGVDYVQKAQTEQADFAVEALAK-----ATYER 431	
QY	87	LVLEMVHHTNASLEK-----LVDTAWEIEMTYAVGECDFEVGKERMLKVKIVK 136	
DB	432	LFRLVHRINKALDRYKOGAFIGILDIAGEFI-----FELNSFQL----- 474	
QY	137	IHPLEKVDDEATEKSDGACDPSDDKENSQIAQDHQKKTETVVKEDGRRSINDRARR 196	
DB	475	-----CINYTEK-----LQQLFNHTMTFLEQEEYREGI----- 504	
QY	197	SPRKUPTSLLKGERKWPKEFLPHKYDVVKLQNEDKIISNVPADSLIRTERPNKEIVRYF 256	
DB	505	-----EW-----NFIDFGLDLQ-----PCIDLI--ERPANPPGVVAL 534	
QY	257	IRHNALRAGTGENAPWVVE-----DELVKKYSLPSKE-----SDFLLDPKYMT 300	
DB	535	L-----DEECFPKATDKTFVEKLVBQGSKSKFKPQLKADFCIIHY----- 580	
QY	301	INPSTKRKNTGSPDRKPS-----KSKTNDSSLSPLN-----PKLWCHVH----- 341	
DB	581	-----AGKVDYKADWLKMNMDPLNDNVATLLHQSSDRFVABLWKDQVDRIVGLQOV 631	
QY	342	--LKKSLSCSPLKVKNSKNSKSPPEHLEEMKWMSP-NKLHTNP-----HLPKGPAPKPKG 395	
DB	632	TGMTETAFGSAKYTKGKMFRTVGQLYKESLTKLMAFLRNTNPNFVRCIIPNH-----EKRA 688	
QY	396	KHSDKPLKAKGRSKGIING-----OKSTGNSKSPKKGLTKPTKMKQMTLLDMAGK 446	
DB	689	KLDPLHLVQLRCNGVLEGIRICRQGFPPRIYFQEPQRYEILTPNAIPK--GFMDGKQA 746	
QY	447	TQMTAPRNSGGTPTTSKPKHPLPAALHLIAYYKENKREDKRSALSCVISKTAR-L 505	
DB	747	CERMIKALELDPNLYRIGQSKIFFRAGVLAHL-----EERDLKITDIIIFQAVCRGC 800	
QY	506	LSSEDLARLPEELRSL--VQKR---YELLEHKRWASMEEQEYKLVKKRELKKLKE 560	
DB	801	LARKAFKKQKQQLSALKYLRNCAAYLKLRLHWQWRVFTKVPLQLQVTRQEEELQAKDEE 860	
QY	561	KAKERREKEMLE-RLEKQKRYEDQELTGKNLPAFLRLVDTPEGILPNTLFGDVAMVVEFLSC 619	

DB	861	LLKVKKEQTKVEGELEEMERKHQQLLEEKNI-----	891
QY	620	YSGLLLPDAQPIITAVSLMEALSADKGGFLYNRVLLVILLQTLLODEIAEDYELGCMKLS 679	
DB	892	-----LAEQLQA-----ETELFAEAEEMPARLAARKQ 918	
QY	680	EIPLTLHSHVSELVRLCLRRSDVQEESEGS---DTDDNKDSAAFEDNEVQ-DE----- 727	
DB	919	ELEELHDL-----ESRVEEEERNQILONEKKKQQAHIQLEQLDEEGARQKL 969	
QY	728	FLEKLETSFFELTSEEKLOIL-----TALCHRIILMTYSVQDHMETRQOMSAEL 776	
DB	970	OLEKV-TAEAKIKWEEBILLEDQNSKFKEKLMEDRIABCSQLAAEESEKAKLAKI 1028	
QY	777	WKERLAVLKEENDKKRAEKOKKENE-AKNKENGKVENGLKTDKTKKIVKPEPQVDTEA 835	
DB	1029	RNKQVMISDLERLKEKTRQELEKAKRLDGET-----TDQDQIAELQAQIDE-- 1080	
QY	836	EDMISAVKSRLLAIQ-AKEREIQ-----EREMKVKLERQAE-----EERIR 877	
DB	1081	-----LKLQAKKEELQALARGDDETLHKNNALKVVRELQAQIAELQEDFE 1128	
QY	878	RHKAAAEKAFQEGIAKAKLVMRRTPIGTDRNHNRYLFSDEVPGLFIEKGWVHDSIDYRF 937	
DB	1129	SEKASRNKAQKQ---KRLD-----SEELEAKTE---LEDTLD--- 1160	
QY	938	NHHKDHVTSGDEDYCPRSKK-----ANLGKN-----ASMTQCHGTATEVAVET 981	
DB	1161	-----TTAAQOEI--RTKREQVAAELKALKEETKNHEAQIDMRQRHATSELSQ 1211	
QY	982	TFP-----KQGNLWFLCDSQKELDELLNCLPQGIRESQLEKRLKRYQDIH 1030	
DB	1212	LEQAKRFANLEKNQKQ-----LETNDKELACEVAVLQVKAESSEHKKKLDAQVE--- 1263	
QY	1031	SIHLARKPNLGLKCDGNQELNLFRLSLIEVATRLQ-----KGLGYVEE 1076	
DB	1264	-----LHAKVSEGR-----LRVELAEKASKLQNELDNVSTLLEAEKKGKFKAD 1309	
QY	1077	TSEFARVISLEKLDGECVIALQASVIKKFQGFMAPKQKRLQSEDSAKTEVDDE 1136	
DB	1310	AASLESQDQTOEL-----LOETROK-----LNLSSRIOLSEKNSLQSOQEE 1355	
QY	1137	KOMVEAKVASALEKWKTAIRE--AQTFSRMHVLGLMDACIKWDSMAENARCKVCPKKG 1194	
DB	1356	-----EAR--KNLEQVIALQSLADTKKVDGDDLTGTIE-----SLEAKKLL-KDA 1401	
QY	1195	E-----DDKILCDCECNKAFHLFCLRPALYVFP--DGEWQCPA-----COPATAR 1237	
DB	1402	EALSQRLEEKALAYDKLEKTKNR--LQQLDLDLTVDLDHQRVASNLKQKQKFDQLLAE 1459	
QY	1238	RNSRGNTYTESASEDSEDESEEE-----EEEEEEEDYEVAGLRP--RKT 1287	
DB	1460	EKSISARYAE--RDRAEAEAREKETKALSARALEALEAKEEFERQNKQLRADMEDL 1516	
QY	1288	IRGHSVIPPAAARSRRPGKPHSTRSOPKAPPVDDAEVDELVLQTKSSRSRQSLQK 1347	
DB	1517	MSKDDV-----GKNVHELEKGRAL-----EQQVEERTQLELED-----ELQA 1557	
QY	1348	CE-----EILHKIVKYSFSPFPREPVTDEAD-----YYDVTIHPMDFTQVQNK 1393	
DB	1558	TEDAKLRLVNNQAKQAQFE---RDLQTRDEQNEEKRLLIQKVRELEAELEDERKQAL 1614	
QY	1394	SCGSYRSVQEFITDMKQVFTNAEVNCRGSHVLSCH----- 1429	
DB	1615	AVASKKME--IDLKLEAQIAEAANKARDEVIKQLKLAQMKDYQRELEEARSRDEI 1671	
QY	1430	-----VKTEOCLVVLHKLPHGPHVVRKPKKPPDRLAEDGEGEPEAVGQSR-- 1477	
DB	1672	FAQSSEKLLKSLAEILQLEELASSERARRHAEQERDELADITNS---ASGKSALL 1728	
QY	1478	DEDRR--GREABIQEWLQDTSLYSAKINSK 1505	
DB	1729	DEKRLRLARIAQLESELEEEQSNMELLNDR 1758	

RESULT 39

D96796
 probable heat shock protein, 53413-59028 [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004
 C:Accession: D96796
 R:Theologias, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; ansen, N.F.; Hughes, B.; Huizar, L.
 Nature 408, 816-820, 2000
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A:Authors: Salzborg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A:Reference number: A86141; MUID:21016719; PMID:11130712
 A:Accession: D96796
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1871 <STO>
 A:Cross-references: UNIPROT:Q9SRD5; GB:AE005173; NID:g6143906; PIDN:AAF04452.1; GSPDB:GN
 C:Genetics:
 A:Gene: F28O16.15
 A:Map position: 1

Query Match 3.3%; Score 264.5; DB 2; Length 1871;

Best Local Similarity 18.3%; Pred. No. 0.00035;

Matches 291; Conservative 223; Mismatches 557; Indels 521; Gaps 65;

QY 33 RTREYEARELYSERIWTCK--STGSSQLTHKEAWEEOEVAELIKKEFPAPWYKLV--- 87
 DB 432 KORQEENDKVGQASEDISLTQLQEGEQFQGGKRDHQENIKELREGQASAEKNIKND 491
 QY 88 VLEWVHNTASLEKLVDTTAWLEIMTKYAVG-EBCDFEVGKERMLKVKVIKHPLEKVDDE 146
 DB 492 ILKPVQKRSEGHKIQKT-FOETNKQPEGYNEKIMETGK-----KINED 535
 QY 147 ATKEKSD---GACDPS-SDKENSSO-----IAQDHQKETTVEKDEGRRESINDRARR 196
 DB 536 GTRKVOEMTRQELDEPASEKERNRSELVKSNTDDEEKEKEIAGTE-RKEKESDR--- 591
 QY 197 SPRKLPTSLKGERKWAPKFLPHKYDVKLQNEDK-----IISNPADSLI---RTERPN 249
 DB 592 -----PKILREQEVADEVAEDTKFPIYGEVKEEIEIAGKEKEFGSD 633
 QY 250 KEIVRYF-----IRHNALRAGTGENAPWVVEDELVKYSLPSKFSDFLLDPYKMTLNP 304
 DB 634 DDIARIVRDTEQLDSNAK---GQE-----EKDMLQELVLEEK---VCDGGKGIIAVAE 681
 QY 305 TKRKTGSPDRKPSKSKTDSLSPLPKLVCHVHLKSLSG-----SPL 351
 DB 682 TKAENKSKRVQETEEQKLKEDT-----CGKHFOKLIEGETSDHGEVEDVEKGGK 732
 QY 352 KVKNSKNSKSPHEHEEM-----MKMSPNK-LHTNFHPKKGPPAKPK 394
 DB 733 RTEAEKRIKDRAREAEIEKDLGVSGRVIKGTIKELVENGIYRNEHEEKDDANRP 792
 QY 395 GXHSDK-----PLKAGRSKGILNGOKS-----TGNSKS----- 423
 DB 793 EKITGITKQELVSLNSQLAQENVEDGDKTQELVEEKIKDCEEEGSESKINTDDVVRKV 852
 QY 424 -----PKGLTPTTKKQMTLLDWAQTKQ----- 449
 DB 853 QGIXEBELYKPKREHGKTITELVEETTGDIYEQEKEETAESDIEACGSLRKVDGIEEHE 912
 QY 450 -----WTRAPRNSG----- 465
 DB 913 LHPEKIHKERNRNVGTAKEPGQEGKEKEKIVSMITTENDNSIDVOETKKERAPGRLE 972
 QY 466 KPHKHLPPAALHLIAYYKENKDRKRSALSCVISKARTLLSSDRARLPEELRSIVQKR 525

RESULT 40

A40997
 myosin heavy chain, striated adductor muscle - scallop (Aequipecten irradians)
 N:Contains: myosin ATPase (EC 3.6.4.1)
 C:Species: Aequipecten irradians

DB 973 SHDKRYKIQELLMEAGHNDRKEEQNENVTAEVELETERVSSKKVQEGKMEDDNS--GKF 1030
 QY 526 YELLEHK--KRWASMSSEORKEVLYKK-----KREELKKLKERAKER 565
 DB 1031 HFEPEKSVEDTHEKREKKVLEBEETYPDKDHTGGGDHNDHKEEQENVIKAEALN 1090
 QY 566 REKE---MLERLEKQ-----KRYEQE-----LTGKNLPAFLVLDTPPEGLPNT 605
 DB 1091 TEEDSPKVEEIEKQDHGELKRSWQAKQETEEKDKTRAMEKNETVEERKQTKDG--- 1146
 QY 606 LFGDVAVVVEFLSCYSGLLLPDPAQPIITAVSLMEALSADKGGFLYLRVVLVILLQTLQD 665
 DB 1147 -----SLGKREGSDPELGG--HERRGESEDRISLVE 1177
 QY 666 EIA-----EDYGLGMLKLSIPLTLHVSSELVRLCLRR---SDVQESSEGSTDD 712
 DB 1178 EISDHKEKVKKDEDYILRSQDTGKVDL-----GERERSKORKIHKSVEDEIGDOEDED 1232
 QY 713 NKDSAAF-----EDNEVODEFLEKLETFEFTFELTSEEKLQILTALCHRI 756
 DB 1233 ABEAAAVSRNENGSSRKVTIEESEKHKQNKIPETSN-PEVNEEDEERVVEK----- 1286
 QY 757 LMTYSVDIMET-----ROOMSAE-LWKERL-----AVLKEEN 788
 DB 1287 -ETKEVEAHVQLEGKTENCKDDGGEERREERGKQGMTAENMLRQRFKTKSDDGIVRKIQ 1345
 QY 789 DKKRAEKOKRKEWAKN-----KNGKVENGKGTDRK--KRIVKFPQVDTEAEDMI 839
 DB 1346 ETKEEPDEBKSSQESSHVVKLVADGSLRNLGFESEKSTVSKMLJLDESKEEHHKI 1405
 QY 840 SAVKSRRLIA--IOAKEREIOEREMKVKLRQ-----ABEERIRKHK 880
 DB 1406 RKPTERSNAPVIEKQGNKNAEEMQDKIDRGNQEIKGQEPYGLVNGEHDKITEYH 1465
 QY 881 AAAEKAFOGGIAKAK-----LVMRTPIDGTDRNNRYWLPSPDEVPGVFIKSGWVHDSI 933
 DB 1466 RGEEXTAENVSTKIQTQKDELEKPRKPSISENHNIFEPMDSOSQDIEBK--SDQA 1523
 QY 934 DYRENNHCKDHTVSGDED-----YCPRSKKANLGNKASMTQHTGTAT 975
 DB 1524 E-KYAKQNKIQEVMNDEKKESYHISERVNEMAKRILQVESKANDGSSKKQNETEGEST 1582
 QY 976 EVAVETTPTKQGNLWFLCDSQKELDELNCLHPOQIRESOL---KERLEKYQDIHSHI 1032
 DB 1583 GLRGR---KKEEN-----HQLVELETSQKGVKDEEVVGAELIIEDEY-DSSRKI 1630
 QY 1033 HUARPNIKGLKSCDGNQELLNPLRSDLEIATRLQKGLGYVEETSEFPAEIVSLF-KLK 1091
 DB 1631 HEHEERMSDKLEMHGEEEMSEKLABEETSDBGSEAKGNRAGKRSRDDGFGVKRKEIVQRK 1690
 QY 1092 DFGECVIALQASVIKKFLOGFWAPKQKRKLOSEDSAKTE--EVDEEKKWVEAKVASAL 1149
 DB 1691 DMDQ-----SFVEKOTSG-----KAKENLNDEEPTKTETKATDNESRKHOIQEQGTS 1738
 QY 1150 EKWKTAIRAEOATFSRMHVLGLMDACIKWDSMAENARCKVCPKGGEDDKILCDECNKAP 1209
 DB 1739 EQ-----ERLKEQRI-----KELVEDRTHCREKEN-- 1765
 QY 1210 HLFCLRPALYEPDGEWQCPACQATARNRSGRNYTEESASE--DSEDESEDEEEEEE 1267
 DB 1766 -----RETEYE--DGS-----SKMIQSIDKEESTEPVDRETSDEDEEELEIE 1805
 QY 1268 EEESEDEEVAGL-----RLRPRTIR 1289
 DB 1806 FEDESEDEAEVIEQTDSNDNDKIRQIRKR 1837

Db 728 HAA-----IVAELEKKEETAKLGEHKEVVLGB-----759
Qy 698 RSDVQESRSGTDDNDKSDSAFEDNEVDQEFLEKLETSEFFELTSBEKLIQILALCHRIIL 757
Db 760 ----QKKEETIILBEKHKDVVTKLGEQHKENIILE-----BEHKDVVTKL-----801
Qy 758 MTSVQDHMETROQMSALWKLAVLKEENDKKRAEKQK-----KEM-EAKKENG 809
Db 802 -----GDQYKEEIAKLKEHVAELEEKHLGEGHEMVDLEKRA 845
Qy 810 KVENGLGKTDKRIKVKFEPQVTEAEDMISAVKSRRLILAIQAKKEREIOE-----R 861
Db 846 DFEGL-EKHKETAKLEGHKSE-----WNEVEKHADPVEGLEEKHAKETAKLGEHR 900
Qy 862 EMVKLERQAE-----ERIRKH-----AAAKAFQEGIAKAKLVNRRTPIGTDNRHNRWLF 915
Db 901 EVVAGLEEKHEVVAELEEKKEIAKLGEHKEVMAELGEHKEVVALEAKHN-----955
Qy 916 SDEVPGLFTKGVWHDSDYRFNHCKDHTVSGDEYDQPRSKKANI/GKNASMTQH-----971
Db 956 -----LEEG--HKEMVAELEKRAHDLVAVLEEQH-----KAEIHK---LGEHKEVV 997
Qy 972 -GTAVAVETTTFKQGNLWFLCDQKDELINCLHPQIGRESQLEKRYQDIH 1030
Db 998 AGIEEKVZAIK-----LAEHKOVTTKLGEQHKEEI-----AKLEDGHKEVYN 1042
Qy 1031 SIHLARKPNLGLKSCDGNQOELLNPL-----RSDLI EVATRLQGLGGLGYVEETSEFEARVIS 1086
Db 1043 EVE-----KKNASLLNLEENHKNEMIKLKEHESASDLVEKLYQKQBEVKN 1090
Qy 1087 -----LEKL-----KDFECVIALQASVKKFLOGFMAPKQRRKLOSEDSAKTEVDEE-- 1136
Db 1091 SNNKIEELTNVINDLSDIMCVKQILEEV-----EKREYNEEINKLIVQNMK 1141
Qy 1137 ----KKWVEE-----AKVASALEKWKTAIREAQTFSRHVLG-----MLDACIKWMSA 1182
Db 1142 DMNDKKILSENEIKLNKLKNYKVFETKNTYKNSEMVNENKERIIVDSVCENTISE 1201
Qy 1183 E-----NARCKVCPKGGEDDKILCDECNKAFHLFLRPAly-----EVPDGEW 1226
Db 1202 SDVEGGGNLKWTLSLKXKERNIFSINDKNESSELVDITKAYINKIEMYYKEIED--- 1258
Qy 1227 QCPACOPATARNRSRGNTY-----ESASESDESDDESEEE-----1266
Db 1259 -----NGKNIEDLNKILDLNLEINLENKNVLTENNKNLKEIEIKDNK 1304
Qy 1267 -EEEREEDYEAGL 1280
Db 1305 LNEKKNTEILNL 1319

RESULT 44
A34736
nestin - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 13-Jul-1990 #sequence_revision 09-Oct-1992 #text_change 09-Jul-2004
R:Accession: A34736
R:Lendahl, U.; Zimmerman, L.B.; McKay, R.D.G.
Cell 60, 585-595, 1990
A:Title: CNS stem cells express a new class of intermediate filament protein.
A:Reference number: A34736; MUID:90150286; PMID:1689217
A:Status: preliminary
A:Molecule type: DNA; mRNA
A:Residues: 1-1805 <LEN>
A:Cross-references: UNIPROT:P21263; GB:M34384; NID:g205663; PID:AA41685.1; PID:g205664
C:Note: the authors translated the codon CGG for residue 168 as Gly, ATC for residue 169
C:Keywords: coiled coil

Query Match 3.3%; Score 263; DB 2; Length 1805;

Best Local Similarity 18.4%; Pred. No. 0.0004;

Matches 325; Conservative 261; Mismatches 601; Indels 580; Gaps 85;

Qy 7 RKPFFLVNPLPGEPPFTTIPHTQE-AFRTREYEALERYSERIWTCKST-----GSSQLTH 62
Db 160 RPPAPPHPRIQ-----PAPEVEDLARRLGEVWRGAVRDYQERVAHWESSLQARLSQ 214
Qy 63 --KEAWEEQEVAEILLKEEFPWYKLVLEWHHTASLEKLVDTAW-----LEIMTYAVG 117
Db 215 AVRGARECRELEVQOQADR-----DSQERREALEQRLGEWRDLQATDKF--- 261
Qy 118 BECDFEVGKMKLVKVIKHPLEKVBDEATKSDGACDSSDDKENSQAQDQHK- 176
Db 262 -----QLAVEALEQEKQGLQSQIAILEGGQQLA-----HLKWS 295
Qy 177 -----ETVVKEDEGRRESINDRARRS-----PRKLPTSLLKGGKRWK-----APP 215
Db 296 LSEVATYRTLLAEANSRLQTPGRGQASGLFLDPKLFENFLGIPEDQVLSVLPAISPT 355
Qy 216 KP---LPHKYDVK-----LQNEDKIINVPADSLIRTERPNKIEIVRYFTRNAL 262
Db 356 SFPSPLPNTLETPTVTAFLKTOBFLQARTPTLASTPIPISEAPCPNABVRAQEVPLSLL 415
Qy 263 RAGTGENAPWVVEDELVKKYSLPK-----FSDPL-----LQPKYM 299
Db 416 QT-----QAP-----EPLWIKATVPSSAILPELEEPGGKQGGHPDDLTSLATNLNPH--- 464
Qy 300 TLNPTKTKNTGSPDRKPKSKSTDNSSLSPNPKLWCHVHLKKSLSGSLPKVKNKNS 359
Db 465 --HPTLEAKDGBSSSRVSSIFQDEG-----QIW---ELVEKEADIEVKVENSQA 511
Qy 360 KSPPEHLE-----EMMQMSPNKLT-----NFHPPKGGPAPK 394
Db 512 KTQESGLDTEETQDSQGLQKETLKALGEEPLMSLKIQNYETAGKENCNSSTEGHGLTLE 571
Qy 395 GHGSKDKPKAKGRSGILNGOKSTGN-----SKSPKGLKTPKTKMQMTLL 441
Db 572 GPEKEQPLKLSLEKNVSEKTELENGVPLSELIGKEDTRTDQELMSPTGLTKRFS-- 629
Qy 442 DMAQGTQWTRAPRNSGGTPTRTSSKPHKLP---PAALHLIAYYKENDREDKRSALSCV 498
Db 630 SLGKESQEVVRPKSEGNLESMTAFKESQHPGLGPGA-----EDQ-----669
Qy 499 ISKTARLLSSEDRA--RUPEE-----LRLSVQRYELLEHKKRWAS-----MSEORKEYL 547
Db 670 --MLERLVEKEDQSPRSPBEEDQACRPLQENQEPGLYBEAQILLETIKESQESL 727
Qy 548 KKKREE-----LKKLKEK-AKERREKEMLERL-----EKQRYEDELTKG-----588
Db 728 RSPDEEDQAGRSLOKENQEPGLYEEADQMLERLETIKESQESLSPENQIRGLPERE 787
Qy 589 NLPAPRLVDTPEGLPNTLFGDVAMVVEFLSCVSGILLPLDPAQVPIPAVSLMEALSADKGF 648
Db 788 NQKSLRYLEENQ-----ETFPVLESRN-----QRPRLSLEVEE-----821
Qy 649 LYLNRVVLLOTLQDEIAEDYGEKMKLSEIPLTLHSVSELVRL-----CLRRS---DV 701
Db 822 ---EQRIKVKLEKVSQDSLGS-----LAEENVQPLRYLEDDDCINKSLLEBDK 865
Qy 702 QEESEGSDDNDKSDAFAF-EDNEVQ-----DEFLEKLE-TSEFFELTSEKLIQI- 748
Db 866 THKSLGSLDRNGDSIIIPQESSETQVSLRPPPEEDQRIVNHLEKESQESFSSSEEEQVM 925
Qy 749 ---LITALCHRIILMTVSVQDHMETROQMSALEWKLAVLKEENDKKRAEKQKREMAKN 805
Db 926 ERSLEGENHESLSVEKED-----QWVESQLEKE-----SQDSKSLSEDSQETFGPLE 974
Qy 806 KENGKVENGLGKTDKRIKVKFEPQVTEAEDMISAVKSRRLILAIQAKK-----854
Db 975 KENAESLRSLAGDQE-----EQKLEQETQOTLRAVGNQMAVSPPEKVPPELPKPLGN 1028
Qy 855 -----EREIOE-----REMKVKLERQAEERIRKHAANAFAFGIACAKLWMRT 901
Db 1029 DOEIARSIGKENQESVLSKEKGIETVKLSLETETIEPLETA-----EDLERRKSIDTQE 1083
Qy 902 FIGTDNRHNRWLFSDVEPGLFIEKGWVHDSIDYRFNHCKDHTVSGDEYDQPRSKKANL 961

Db 1084 P-----LWSTEV-----AETVEPDEPP----- 1103
QY 962 GKNASNTQHGTEVAVETTPKQGNL-----WFLCDSQKELDELNCLH-POGIRES 1015
Db 1104 GSLGSVDENRETLTSL-----EKESQELSSLGK-----NVETRVDSQOCLQVEEGLQBE 1154
QY 1016 QUKERL-----EKRYQDIH-----STHLARKPNLGLKSCDGNQE 1050
Db 1155 QHQSRELYKQELPSSGNOQRWEDVVEGKAVGQAPLATTGVGTEDKAEHLRGGGEBE 1214
QY 1051 LL--NFLRSLDLVATRLQKGLGYVEETSEFEARVISLEKLDGFCVIALQAQSVIKKF 1108
Db 1215 AAABGELLQDI VGEAWSLGS-----SEPKQORVPAEALONLEGG-----AUEVPVQSM 1263
QY 1109 LOGFMAPKORRKLOSEDSAKTEVD--BEKQWVEEAKVASALEKWKTAIREAQTFSRMH 1166
Db 1264 -----PEVTER--DEDRAGAEQDSIEVTGLGEAARTGLEQEVVWGLEDPHFAR-- 1312
QY 1167 VILGLMDACIKWDMNAENARKVC-----PKKGEDDKLILCDECNKAFHLFCLRALPYEVPD 1223
Db 1313 -----EEAIPPSLGEESYKAKIAQGLEPGKEPK-----EAGALDSGILELPK 1355
QY 1224 GEMQCPACOPATARRNSRGRNVTESASDSDDE--SDSEEEEBEEDVEVA----- 1278
Db 1356 TSEALECQCHRESSMEG--WEEBEASLETSDHGSDAPQRPPEDEDEGAQAALTAP 1413
QY 1279 GLRL-----RPRKTIKGHSVIPPA-----ARSGRRPGKKHSTRRSQKAPPVDDAEVDLIV 1331
Db 1414 GPKLLEPCSPILITDAHELQQAEGIQEAGWQPEAGSEALERVE-NEPEFGLGRIPE-G 1471
QY 1332 LQTKSSRRQSLQELQKCEILHKIVKYSWFRPV-----TRDEADYDV----- 1379
Db 1472 LQDWEEGREES-EADDLGETL-----PDSTPLGLYLRSASPCKMDLAGEQRLSPQ 1520
QY 1380 -----ITHPMDFOIVQNKCSGVSRSVQEFLLTDMKOVFTNAEVVNCRG 1422
Db 1521 GQAGKEDWGPVPAQAQGLSGPPEEBEQH-----GSDLSEEP-----EDLGTBASL----- 1568
QY 1423 SHVLSQWVTEQCLVLLHKLPHGHPYVRRKRKFPDLAE-----DEG----- 1466
Db 1569 -----LPGVP-----KEVADHVGPVPPVLPQACWQOGGESDGF 1601
QY 1467 --DSEPEAVQSRDEDRRSREAEIQEW 1491
Db 1602 ADEESGEEGBEDADEGAESGAQWW 1628

RESULT 45
A47371
transcription initiation factor IID 230K chain - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C:Date: 16-Feb-1994 #sequence_revision 18-Nov-1994 #text_change 15-Mar-2004
C:Accession: A47371
R:Kokubo, T.; Gong, D.W.; Yamashita, S.; Horikoshi, M.; Roeder, R.G.; Nakatani, Y.
Genes Dev. 7, 1033-1046, 1993
A:Title: Drosophila 230-kD TFIID subunit, a functional homolog of the human cell cycle g
A:Reference number: A47371; MUID:93279463; PMID:8504928
A:Accession: A47371
A:Status: preliminary
A:Molecule type: mRNA; protein
A:Residues: 1-2068 <KOK>
A:Cross-references: GB:S61883; NID:g385550; PIDN:AAB26991.1; PID:g385551
A>Note: sequence inconsistent with nucleotide translation
C:Note: sequence extracted from NCBI backbone (NCBIN:133002, NCBI:P133003)
C:Genetics:
A:Gene: FlyBase:Taf250
A:Cross-references: FlyBase:FBgn0010355
A:Superfamily: transcription initiation factor TFIID subunit 1 (TAF1), animal type; brom
C:Keywords: transcription initiation
F:1498-1553/Domain: bromodomain homology <BRO1>
F:1620-1675/Domain: bromodomain homology <BRO>

Query Match 3.3%; Score 262.5; DB 2; Length 2068;
Best Local Similarity 19.8%; Pred. No. 0.00049;
Matches 315; Conservative 216; Mismatches 592; Indels 471; Gaps 77;
QY 107 WLEIMTKYAVGEBCDFEVGKEKMLKVIKIHLPLEKVDDEATEKKS DGACDSSPSDDKENS 166
Db 361 WYDILEVDSGE--GFNYG---FKTKAASISSQQLKDERRVKSPEDDVEDPSIADDAF 414
QY 167 SQIAQHQKQKTVKVEDGRRESINDRARRSPKLPSTLKKGERKWAAPK-----FL 218
Db 415 LMVSQLHWEDDVW---DG-----NDIKAKVLQKL--NSKTNAAAGWLPSSGSRTAGASQ 464
QY 219 PHKYDVKL-----QNEDEKIIISNPADSLIRTERPPNKEIVRYEIRNA 261
Db 465 PKGFSMPVSGSSSKOGSGASSKKAQOAAKPAEAPDDTWTYSLFPVNEELIY----- 518
QY 262 LRAGTGENAPVWVEDLVKYSLSKFSDFLLDPYKMTLPNPSTPKRKTGSPDR-KPSKK 320
Db 519 -----KW--EDEVIWDAQVSK-----VPRFKVLTDPDENIILGIPDDIDPSK- 561
QY 321 SKTDNSSLSPLNPKLWC-HVHLKSLSGSPLKVNKSNKNS-----KSPBEB 365
Db 562 ---INKSTGPP--PKIKIPHPHVKS-----KILLGAGVINVLAEPTPPPPKSPDR- 609
QY 366 LEEMKMWSP-NKLTNPHIPKGGPPAKKPGKHSKDKPLKAKGRSGILNGOKSTGNSKSP 424
Db 610 -----DPTNISNDTYTTPKTEPLR-----LKVGNLI-----QHST-----P 642
QY 425 KKGLKTP--KTKMKQMTLLDMAKGTQKMTAP--RNSGGTPTRTSSKPKHKLPPAALHLIA 480
Db 643 VVELRAPFVPTMGPMNV-----RAFRPPLKYSHG-PMAQSIHPVFP--LLKTA 692
QY 481 YYKENKOREDKRSAL---SCVISKTARLLSSED-----PARLPEELRSIVQRYELLEHKK 533
Db 693 ---KKAQKVERIASGGGVFMRNPEDLSGRDGDIVLAEFCEEHPPLI----- 739
QY 534 RNASSEQRKEYLKKKEELKKLKEKAKERREKEMLERLEKOKRYEDQELTGKLPAP 593
Db 740 -----NOVGMCSKIKNYKKAEND----- 759
QY 594 RLVDTPTEGLPNTLFGDVAMVVEFLSCYSGLLLPDPAQYPIITAVSLMEALSADKGGF----- 648
Db 760 -----SGPDYVYGEVAF--HTSPFLGILHPG-----QCIAIENNMYRAPI 800
QY 649 ---LYLNRLVILLQTLQDEIAEDYGEKMK--LSEIP-----LTHSVSELV- 692
Db 801 YPHKMAHNDFLVIRTRNNYWIRSVNSIYTVGOECPLVEYVPGPNSKRANNTRDFLQVFIY 860
QY 693 RLC-----LRSDVOE---ESEGSDTDDNKDSAAFEDNEVQDS-FLEKLETSEFF 738
Db 861 RLFWKSRDNPRIIRMDLIKQAPPAHSESIKRLKQCADFKRTGMDSNWVJIKP-----F 916
QY 739 ELTSEBKILQILTA---LCHRIIMLYSVQDHMETROQMSAELMKERLAVLKEENDKKRAEK 795
Db 917 RLPSEEEIRAMVSPQCCAYFSM-----IAAEQLKADAGYGEKFLFAPQEDDDDEEQL 969
QY 796 QKRKMEAKNKENGKVENGLGKTRDKRIVKPEPQVDTAEADMISAVKSRLLAI----- 850
Db 970 KLDDE-----VKVAPMNTTRA---YIQAMRGKCLLQJSGPAD 1003
QY 851 -----QAKKERIOEREMKV-----KLEQAEERIRKHAARAK 885
Db 1004 PTGCGEGSYVRVKNKPTQTKEEQSQPKRSVTGTDADLRRLPLQRAKELLQFKVPEB- 1062
QY 886 AFQEGIAKAKLVMRRTPIGTDNRNHNRYWLFSDVEVPLFIKRGWVHDSIDYRFNHHCKOHT 945
Db 1063 -----EIKKLSR-----NEVIDVRTLSTERAKACGEGMDKFSRGNRESI 1102
QY 946 VSGDEDY---CPR-----SKKANLGNKASMNTOHGATATEVAVETTTKQOQLWFLCDSQKE 999
Db 1103 AEHQRYKEECORIPDLQNRVLASSEVLSTDEAESS-ASEESDLELGNLMLSNKKT 1161
QY 1000 LDELNLCLHPQGIRESQLEKRYQDI IHSI---HLARKPNLGLKSCDG-----NOELL 1052

Db 1162 STQL-----SRREELER--QELLRLDDEHGGPSSGGAGKAGKDDPQQQML 1208
QY 1053 NFLRSLIEVATLQGGGLGYVEETSEFEARVLSLEKLDGFCVIALOASVILKKFLOGF 1112
Db 1209 ATNQCRIILRITFRNDG-----KEY-TRVTVRR-QPVIDAYIKIRTKDEQIFKQF 1261
QY 1113 MAPQKRRKLQSDSAAKTEVDEEKKWEEAKVASALEKWKTAIRBAQTFSRHHVLLGML 1172
Db 1262 -----ATLDEQKKEKREKRIQEQ--LRRIKRQERERLAQLAQKQLPQGM 1309
QY 1173 DACIKWMSA-----ENARCKVCPKGGEDDKLLDCBCKAFHLFCRLPALYVDPG 1224
Db 1310 PTLGDPKSSGGSHKERSGYKEVFSR-----KKFKL-----KPDLL 1347
QY 1225 EMCPAC-OPATARRNSGRNY-----TEESASEDSDDSEDEEEEEEEDVEVAG 1279
Db 1348 KLCGAGCGVGHRTWKACPLYSGMQSSLSQSPSLADDFDESEKEMTWDDDLVNVVG 1407
QY 1280 LRLRPRTIRGKHSVIPPAAARSRRPG---KKPHSTRRSQPKAPPVDDAEVDELVLQTKR 1336
Db 1408 TKVTSSKILKRGGDDGKRRSSGSGFTLKVPRDA-MGKKKRVGGDLHCDYLQRHNT 1466
QY 1337 SSRQS---LELOKCEILHKIVKRFPS-WPPREPVRDEAEYDYVITHPMDFTQVQNK 1392
Db 1467 ANRRRTDPVVVLSLLEIIHNEILRSPDVSFPLFPVSAKKVPDYRVVTKPMDLQTMREY 1526
QY 1393 CSCGSYRSVOERLTDMKQVFTNAEVN-CRGSHVLSQVTKTEOCLVVL----- 1439
Db 1527 IRORRYTSRMFLIEDLQIVDNLIIYNGPQSAVTLAAQRMFSCFELLAERDKMLREK 1586
QY 1440 -----LH---KHLF-GHPYVRR-KRKKPPDLAEDGSEPEAVG 1474
Db 1587 AINPLDDDDQVALSFIDKLHSQIKQLPESWFLPVPVKQVQKDYVYVTKRPMDLTIG 1646
QY 1475 QSRDEDRRREABIEQLQDTSIYSAKINSKDN 1508
Db 1647 KNIEAHRYHSRA--EVLADIELIAT--NCEQYN 1675

RESULT 46
S21801
myosin heavy chain, neuronal [similarity] - rat
N:Alternate names: myosin II
N:Contains: myosin ATPase (EC 3.6.4.1)
C:Species: Rattus norvegicus (Norway rat)
C>Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 09-Jul-2004
C:Accession: S21801; P00013; S18134
R:Sun, W.; Chantler, P.D.
J. Mol. Biol. 224, 1185-1193, 1992
A:Title: Cloning of the cDNA encoding a neuronal myosin heavy chain from mammalian brain
A:Reference number: S21801; MUID:92235856; PMID:1569576
A:Molecule type: mRNA
A:Residues: 1-1999 <SUN>
A:Cross-references: UNIPROT:Q63731; EMBL:X62659
R:Sun, W.; Chantler, P.D.
Biochem. Biophys. Res. Commun. 175, 244-249, 1991
A:Title: A unique cellular myosin II exhibiting differential expression in the cerebral
A:Reference number: P00013; MUID:91151356; PMID:1998509
A:Accession: P00013
A:Molecule type: mRNA
A:Residues: 1914-1998, 'I' <SU2>
A:Experimental source: brain
C:Superfamily: myosin heavy chain; myosin motor domain homology
C:Keywords: actin binding; ATP; coiled coil; hydrolase; methylated amino acid; nucleotide
F:84-763/Domain: myosin motor domain homology <MMOT>
F:174-181/Region: nucleotide-binding motif A (P-loop)
F:541-575/Region: actin binding #status predicted
F:653-675/Region: actin binding #status predicted
F:836-1999/Domain: actin binding #status predicted
F:836-1276/Region: S2
F:1277-1999/Region: light meromyosin

F:125/Modified site: NG,NG,N6-trimethyllysine (Lys) #status predicted
F:180/Binding site: ATP (Lys) #status predicted
F:693,703/Active site: Cys #status predicted
F:1916/Binding site: phosphate (Ser) (covalent) #status predicted
F:1943/Binding site: phosphate (Ser) (covalent) #status predicted

Query Match 3.3%; Score 261.5; DB 1; Length 1999;
Best Local Similarity 19.4%; Pred. No. 0.00052;
Matches 241; Conservative 230; Mismatches 462; Indels 311; Gaps 56;

QY 75 LLKEFPWAYEKL-VLEVMHNTASLEKLVDTAWLEIMTK-----YAVEECDFEYVGEK 128
Db 793 LARKAPAKQOQSALKILQNRCAAYLKLRHWQMRVFTKVKPLQLQVTRQEBELQAKDE 852
QY 129 MLKVIVK-----IHPLEKVDDEATEKKSDGA----- 155
Db 853 LMK-KVEKQTKVYAELEEMERKHQQLLEKNILAQLOAETELFAEAEMRARLAAKQE 911
QY 156 -----CDSPSDKENSQIAQDHQKQKTVVYKED--EGRRSINDRARRSRPKLPTSJK 206
Db 912 LBEILLHDLESRVVEEERNOILQNEKKQGHKNDLEEQLEDWESARQKLEKVTTEAK 971
QY 207 KGERKWAAPPKFLPHKYDVKLQNEKLIISNPAD--SLIRTPPPNKEIVYFIRHNALRA 264
Db 972 L--KKLEEQIILEDQNCCKLAKEKLLIEDRIAFTNLTEEBEKSSTAKLKNKHEAMIT 1029
QY 265 GTGENAPWVVEDLVKYSLSPSKSFDFLDDPYKYMTLPSTKRXNTGSPDRKPSKSKTD 324
Db 1030 DLER-----LRREKQRELEKTRKLEGDSTD 1058
QY 325 NSSLSPLNPKLCHVHLKLSGSPKVKNSKNSKSPHEHLEEMMMKMSPNKLIHTNPHI 384
Db 1059 LSDQIAELQAQI-----AELKMQLAKK---EEELQAALARVEEBAQKQNAL 1102
QY 385 PKGPPAKPGKHSKPKLAKGRSGKILNGQKSTGNSKSPKGLKTPKTKMKTMLDMA 444
Db 1103 -KKIRELESQISELQEDLESERANKAEKQK-----RDLGELEALKTELEDLT--DST 1154
QY 445 KGTQKMTAPRNS-----GGTPRTSSKPHKHLPPAALHLIAYYKKNKDEDRK-SALSCV 498
Db 1155 AAQQL-RSKRQEVNLIKLTLEEAKEHE-----AQIQMRQKHSQAVELEAQ 1203
QY 499 ISKTARLLSSDRARLPPEELSLVKRYELLEHHK---RWASMSERQKRYLKKREELK 555
Db 1204 LQTKRKVANLEKAK-----QTLNENGERELANEVVLQGGRDSEHKKK-VEAQQLQELQ 1257
QY 556 KKLKAKKERRKEWMLERLEKOKRYEDQELTGKULPAFLVDTPEGLPTLPGDVAMVVE 615
Db 1258 VKFNE--GERRVTELADKVKTK-LQVELDNVTG-----LLSQSDSKSKLTKDFSALES 1307
QY 616 FLSCYSGLLLPDAQYPIITAVSLWEALSADKGGFLYLNRLVILLQTLLODEAEDYGLG 675
Db 1308 QLQDTQELLQENRQKLSLSTKQLQVDEKNSF-----REQLEEEEEEAK 1352
QY 676 MKLSIPTLTHS-VSELVRL-----CLRRSD-----VOEESGSDTDNDKSAAPEDNE 723
Db 1353 HNLEKQIATLHAQVADMKKMEDSVGCLTAEEVKKLQKDLGLSQRHEEKVAAVDK-- 1410
QY 724 VODEFLEKLETSEFFELTSEEKLIQILTALCHIRILMTYSVQDHMETRQMSAE--LWKERL 781
Db 1411 -----LEKTKR-----LQQLDLDLVDLHORQACNLKQKPKPQLLLEAITSAKY 1460
QY 782 AVLKSENDKKRAEQ-----KKKMEA-----KNKE-----NGKV----- 811
Db 1461 AEERADAEERAEKATKELSLARAELEAMEQAEFLRKVLQEMTERLDEMSKVDKAKSV 1520
QY 812 -ENGLGKTKDKRKIRIVKFEQVQVTEADMTISAVKSR--LLATQAKKEREIOBR--- 861
Db 1521 LEHEKSKLGRQOVWEEKTQL-LESEDELAQTEDAKLRLEVNQAMKAQFEDLDQGRQDD 1579
QY 862 --EMKVKLQKQ-----ABEERIRKHAAKAEKAFQEGIAKAKLVMRRTPTGTD-RHNRYW 913
Db 1580 SEEKQKLVQVREMEAELEDDQKEMSRARA--VKKLEMDLKDLEAHIDSANKNR-- 1633

QY 914 LFSDEVPGLFIEKGWHDSDYR-FNHCKOHTVSGDEDCPRSK-----KANLGKNASM 967
 Db 1634 -----KIQURNQAQMKQCMRELDTRASREIALQAQXENKLLKSM 1678
 QY 968 NTQOH-GTATEVAVETTTTPKQGNLWFLCDSQKELDELNCLHPQGIRESQKLEREKRYQ 1026
 Db 1679 EAEMTQLOEELAAERAKQAQ-----ERDELADEISNASGKAGLAKELR-RLEARIA 1732
 QY 1027 DIHSHILARKNLGKSCDGNQELN-----FIRSDLIEVATRLQKGLGYVE----- 1075
 Db 1733 QLEEE-----LEEEQNGTELNDRLLKKNLQIDQINADLNLERGHAQKNENARQQ 1782
 QY 1076 ---ETSEPEARVISL---KLXDFGECVIALQASVKKFLOQFMAPKQKRLQSEDSAK 1129
 Db 1783 LERQNKELVKLOEMEGTVKSK-YKASITALEAKTAQ---LEEQDNETKERQAACKQVRR 1839
 QY 1130 TE-----EYDEKQWEE-----AKVASALEKWKTAIRAQAQ 1160
 Db 1840 TEKLLQVLLQVDDERRNAEQYKDQADKASTRLKQLKQLEAE 1883

RESULT 47
 myosin heavy chain beta, cardiac muscle [similarity] - golden hamster
 C:Species: Mesocricetus auratus (golden hamster)
 C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 09-Jul-2004
 C:Accession: I48153; A28298
 R:Wang, R.; Sole, M.J.; Cukerman, E.; Liew, C.C.
 J. Mol. Cell. Cardiol. 26, 1155-1165, 1994
 A:Title: Characterization and nucleotide sequence of the cardiac alpha-myosin heavy chain
 A:Reference number: I48153; MUID:95115033; PMID:7815459
 A:Accession: I48153
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1934 <RES>
 A:Cross-references: UNIPROT:P13540; GB:I12104; NID:G402371; PIDN:AAA62313.1; PID:G402372
 536, L, 1538-1555, K, 1557-1594 <JAN>
 R:Jandreski, M.A.; Sole, M.J.; Liew, C.C.
 Nucleic Acids Res. 16, 4737, 1988
 A:Title: Sequence of cDNA encoding the Syrian hamster cardiac beta-myosin heavy chain.
 A:Reference number: A28298; MUID:88247788; PMID:3380703
 A:Note: the authors translated the codon GTG for residue 1504 as Leu
 C:Genetics:
 A:Introns: 66/3; 114/3; 167/1; 176/2; 212/3; 243/3; 265/1; 298/1; 332/3; 379/1; 418/3; 423/3; 1389/2; 1450/3; 1506/1; 1547/3; 1650/3; 1718/3; 1760/3; 1852/3; 1884/3; 1929/3
 C:Superfamily: myosin heavy chain; myosin motor domain homology
 C:Keywords: actin binding; ATP; cardiac muscle; coiled coil; heart; muscle; nucleotide binding site; myosin motor domain homology <MOT>
 F:177-184/Region: nucleotide-binding motif A (p-loop)

Query Match 3.3%; Score 259.5; DB 2; Length 1934;
 Best Local Similarity 18.4%; Pred. No. 0.00062;
 Matches 288; Conservative 236; Mismatches 460; Indels 579; Gaps 72;

QY 249 NKEIVRYFTRHNLALRAGTGENAPWVVEDELVKYSLSPSKFSDFLD-----PYKY 298
 Db 481 NEKLOQFFNH-----MFVLEQEEYKKEGIEWTFIDFGMDLQACIDLIEKPMRI 529
 QY 299 MTL-----NPSTKKNVGSFDRKPSK-----SKTDNSS 327
 Db 530 MSILBECMFPPKATDWTFRKALVDNHLGKSNFQPRNVKQKQEAHFLSVHYAGTVDYNI 589
 QY 328 L-----SSPLNP-----KLWCHVHLKSLSGSPKLVKSKSKSPKEHLEEMM 370
 Db 590 LGWLQKNKDPLETNVGLYQKSLKLLSNLFANYAGADAPVDKGGKAKKSSSFQTVSVL 649
 QY 371 KMSPNKLTNPH-----IPKGGPPAKKPGHSDKPLKAKGRSKGILNGOKSTGN 420

Db 650 HRENKLMTNLSTRTHPHFVRCIIPNE---TKSPGVMDNPLVMHQLRCNGVLGIRIC-- 704
 QY 421 SKSPKGLKT-----PKYOKQMTLLDMAKGTQKMTAPRPNSSGTPRTSK 466
 Db 705 ---RKGPNNRILYDFFQRYRIINPAAPPEGQFIDSRKGAEL-----LSSL 748
 QY 467 PHKLPPAALHLIAYYKEN-----KDRDKSALSVCISKTARLSSSDRARLP-ELRS 520
 Db 749 DIDHNOYKFGTKVFFKAGLGLLEEMDER--LSRIITR-----IQASRGLLSRMEFK 802
 QY 521 LVQRYEYLL-----EHKRWASMSSE--ORKEYL 547
 Db 803 LLERRDSLVIQWNIIRAPMGVKNWPMKLYFKIKPLKLSAETEKEMATKSEFGVKDAL 862
 QY 548 KK---KREELKK---LKEK-----AKER-----REKWL 571
 Db 863 EKSEARRKELEBKMSVLSLQEQNDLQVQABQDNLADAEERCDQLIKNKIQLEAKVKMT 922
 QY 572 ERLEKOKRYEDELGTGNLPAFLPRLVDTPEGLPNTLFGDVAMVVEFLSCVSGLLLPDAQVP 631
 Db 923 ERLEDEEM-NAELTAK----- 938
 QY 632 ITAVSLMEALSADKGGFLYLNRLVILQTLLOEIAE---DYBELGMKLSIPITLHSV 688
 Db 939 -----KRLDECSSELKRDIDDLTLLAKVEXDKIAT 970
 QY 689 SELVRLCRRSDVQSESGSD---TDDNKDSAAPEQNEVQDFLEKLETSSEFELTSEK 745
 Db 971 ENKV-----NLTEEMAGLDEIIAKLTKEKALQ--BAHQOALDLO-----AEDBK 1015
 QY 746 LQILTALCHRLTMTVSQD-----HMETROQMSABLWKERL-----AVLKEENDKK 791
 Db 1016 VNTLTK---SKVLEQOQVDDLEGSLEQEKVVRMDLERAKRLEGLDKLTQESIMDLNDKQ 1073
 QY 792 RA-EKQKREMEAKNKENGKVEN--GLGKTDRK-----RIVKPEQVDEAEADISAV 842
 Db 1074 QLDEKLLKKDPFL-NALNARIEDEQALGSLQKLLKELQARIEELEEELEA---RTARA 1129
 QY 843 KSRLLALQAKKERIQR-----EMKVKLERQAESERIRK-----HKAAA- 883
 Db 1130 KVEKLSRLSRELEISERLEEAGGATSVQEMNKREAEFQKMRDLLEEATLQHEATAA 1189
 QY 884 ---EKAFOGIAK-AKLV--MERTPIGTDRNHNRYLWFSDEVPGLF-----IEKGMV 929
 Db 1190 ALRKHHADSVAELGEIDINLQVRQKLEKSEFKLEDDVTSNMEIILKAKANLEK--M 1247
 QY 930 HSDIDYRFNHHCQHTVSGDEDCPRSKKANLGKNASMTQHTGATEVAVETTTPKQGN 989
 Db 1248 CRTLEDQWNEH-----RSKABETQSVNDLTSQAKLQTENG----- 1284
 QY 990 LWFLCDSOKELDE---LLNCLHPQGIRESQ---LKERLE---KRYQDIHSHILARKPN 1039
 Db 1285 -----ELSRQDLDEKALISQLTRGKLTYYQOEDLKRLQLEEEVKAKNTLAHALQASR--- 1336
 QY 1040 LGLKSCDGNQELNLFRLSDLIEVATRLQKGLGYVEETSEFEARVISLEKLDKDFECVIA 1099
 Db 1337 ---HDCC-----LLRQ-----YEEET-EAKAEL-----QCULS 1361
 QY 1100 LQASVKKFLQGMFAPKQKRRKQSEDSAKTEEDVEDEKQW-----EAKVA-----S 1147
 Db 1362 KANSEV-----AQWRTKYETAIDAIQRTTEELEAKKLAQRLQDAEAEVAVNAKS 1411
 QY 1148 ALEKWKTAIREAQTSRMHVLGMDACIKWDMSEANARCKVCPKKGEDDKLILCECNK 1207
 Db 1412 SLEKTK-----HRLQNEIED--LWVDVERSNAALDKQRNFDKILAEWKOK 1458
 QY 1208 -----AFHLFCLRPAL-----YEVDPGEWQCPAC 1231
 Db 1459 YEESQSELESQKARSISTELFKLNAYESLEHLETFKRNKNLQSEISD----- 1510
 QY 1232 QPATARRNSGRNYTE-----ESASEDSEDESDEE-----E 1263
 Db 1511 --LTELQGSTGKSIHELEKIRKQLEAKENKMLQSALEAEASLEHEEGNITLRAQLBFOIK 1568

Qy	1264	EEEEEEEDYEVAGRLRPRKTIRGKSHVPPAARSGRRPGKKPHSTRSQKAPPVD	1323
Db	1569	AEIERKLAEKDEEMEQAKRNHLRVVDSLQTSLSAETR-----RNEALRVKKKM	1617
Qy	1324	DAEVDLVLQTKRSSRRSLOLKEELHKIVKYRESWFPREPVTDEADYDVITHP	1383
Db	1618	EGDLNEMEIQLSHAN-EMAAEAQKQVKSLOSLLK-----DTQIQDDAVRANDDLKEN	1669
Qy	1384	MDFTQVQNKSCGSGYSRVQSFPLTDMKQVFTNAEYVNCRGSHVLSCHWVKTEQCILVLLHKH	1443
Db	1670	IAIVERNNL-----LQAELELRVAVQTERSRLAEQEL---IETSE-RVQLLHSQ	1718
Qy	1444	LPGHPYVRRKRKFPDRLAEDGDESEPAVQCSDEDRRSRE-----AEIQEWLQDT	1495
Db	1719	--NTSLINQKKWDADLSQLQTEVE-EAVQECRNAEAKKAITDAAMWAEELKKEQDT	1774
Qy	1496	SLY 1498	
Db	1775	SAH 1777	

RESULT 48

Ts1505

hypothetical protein F5E19 70 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 18-Aug-2000 #sequence_revision 18-Aug-2000 #text_change 09-Jul-2004

C:Accession: Ts1505

R:Sato, S.; Nakamura, Y.; Kaneko, T.; Kato, T.; Asamizu, E.; Kotani, H.; Tabata submitted to the Protein Sequence Database, August 2000

A:Reference number: Z25394

A:Accession: Ts1505

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-853 <SAT>

A:Cross-references: UNIPROT:Q9LFE4; EMBL:AL391147

A:Experimental source: cultivar Columbia; BAC clone F5E19

C:Genetics:

A:Map position: 5

A:Introns: 6/2; 79/3

A:Note: F5E19 70

Qy	802	EAKKNGKVGELGKTDRKRIVKPEPQVDTEADMI SAVKSRRLLAIQ---AKKERI	858
Db	355	EGSNDKLHDTETEI--TDLKRIVTLTETTVAKQKEDL--EVSEQLRSVEEVS	410
Qy	859	-----QREMKVKLEROA-----	878
Db	411	EKLKSELTVEEKNRALKKEQDATSRVQRLSEKSKLLSDLESSKEEKS	470
Qy	879	---HKAAA-----EKAFQBG-----TAKAKLVMRRTPIGTRDHNHR	922
Db	471	SALHEVSSSEGRELKEKLLSGDHEYETQIDDLKLVIKAT-----NEKYEN	519
Qy	923	FIEKGWVHDSIDYRPN--HHCKDHTVSGDDEYCPRSKKANL-----	966
Db	520	-----ARHEIDVLVSAVEQTKGHFESSKKDW--EMKEANLVNYVKMEED	571
Qy	967	MNQHGTTATVAVETPTTKQGNLWFLCDSQKELDE-----LLNCLHPQGI	1021
Db	572	RDLNLLKRTTEEAADAAWKKEAQTK-----DSLKEVEEIVYLQETLGE	627
Qy	1022	--EKRYQDIIH-SIHLARKPNLGLKSCDGNQOELLNFRSDLI	1078
Db	628	DKETEFQNVJHNEEDLKAQEDVSLKKI-----EELSKLLEAAILAKQ	679
Qy	1079	EFEARVISLEKLKDFGECVIALQASVIKKFLQGFMAPKQKRRKLQSEDS	1138
Db	680	ESEKQYDLLPKVVEFS-----SENGHRSVEEKS	717
Qy	1139	MVEAKVASALEKWKTAIRAEOFTSRMHVLLGMLDACIKWDSAEANARCK	1198
Db	718	--PQEQISNGSNGN-----GMEEKEVNGKPEVETKEKKEKQDESQ	759
Qy	1199	LILCDECNKAFHLFCLRPALPYEVPDGEWQCPACQATARRNSGRNRYTE	1258
Db	760	---DDSVVEVIFQM-----WESQCEKKEKAFDPKKSELESSEEDSS	802
Qy	1259	SDSEEEEEEEEEE---EEDYEYAGLRLRPRKTIRGK	1291
Db	803	SDKTSTENIDETGNALTAEDQLTWEEKIKKKKTLGK	839
RESULT 49			
A42091			
transcription activator SNF2/SWI2 homolog brm - fruit fly (Drosophila melanogaster)			
C:Species: Drosophila melanogaster			
C>Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 09-Jul-2004			
C:Accession: A42091			
R:Tamkun, J.W.; Deuring, R.; Scott, M.P.; Kissinger, M.; Pattatucci, A.M.; Kauff			
Cell 68, 561-572, 1992			
A>Title: brhma: a regulator of Drosophila homeotic genes structurally related			
A:Reference number: A42091; PMID:92154670; PMID:1346755			
A:Accession: A42091			
A>Status: preliminary			
A:Molecule type: mRNA			
A:Residues: 1-1638 <TM>			
A:Cross-references: UNIPROT:P25439; GB:M85049; NID:g157011; PIDW:AAA19661.1; P			
A:Experimental source: iso-1			
A>Note: sequence extracted from NCBI backbone (NCBIN:82354, NCBIP:82360)			
C:Genetics:			
A:Gene: FlyBase:brm			
A:Cross-references: FlyBase:FBgn0000212			
C:Superfamily: human SNF2alpha protein; bromodomain homology			
C:Keywords: transcription regulation			
F:1451-1506/Domain: bromodomain homology <BRO>			

Query Match 3.3%; Score 259; DB 2; Length 1638;
Best Local Similarity 19.3%; Pred. No. 0.00053;
Matches 283; Conservative 198; Mismatches 428; Indels 568; Gaps 69;

Qy 389 PPAKPKGKSHDKPLKAKGRSSKGILNQCKTSGNSKSPKKGLTKPTKTOKWQTLLDMDAGTQ 448
 | | | | | | | | | | | | | |
Db 319 PPPPEPHQH-QLPNGGKPLSMGSGGGQPIIPSS-----PMQPVQRVTLFGMPPPGSQ 368

Db 832 LARAFAKQOQI SALKILORNCAYILKLRHQRVFTKVKPELLQVTRQEEBLQAKDEE 891
Qy 129 MLVKV-----IVKIHPLKVDDEATEKKS-----DGACDSSPSDKENSQIAOQHKK 176
Db 892 LMKVKEKOTVKEAELEBEMERKHQOLLEKNILAEQLQAETELFAEAEEMRARLAACKQEL 951
Qy 177 ETVVKEDEGRESINDBARRSPKRLPTSLKKGKRWAPPKFLPHKVDVKLQNEDK-----231
Db 952 EEILHLESVEEEER-----NQILQNEKK-----KMOGHIQDLEEQDDEEGARQ 998
Qy 232 --IISNVPADSLIRTERPPNKEIVRYFIRHNALRAGTGENAPWVVEDELVKKYSLSPKFS 289
Db 999 KLOLEKVTAEAKJK-----MEEIILLLEDQNSKF- 1028
Qy 290 DFLLDPYKYMTLNPSTKRKNTGSPDRPKSPKSDTNSLSPLNPKLWCHVHLKSLGSGS 349
Db 1029 -----LKEKKLMEDRIA-----ECTSQLAEE 1049
Qy 350 PLKVKNSKNSKSPKE-----HLEEMKQMSPNKLTHTFHPKGGPPAKPKGSHDKPLKAK 405
Db 1050 EEKAKNLAKLNKQEMWITDLEERL-----KBEKTRQELEKAK 1088
Qy 406 GRSGK-ILNGOKSTGNSKPKGLTKPTKQKQMTLLDMAKGQTKVTRAPRNSGGTPRFS 464
Db 1089 RKLDETTDLQDQIAELOQIAELKIQLAKKEEELQAAALARGDEAVQ-----KNN 1140
Qy 465 SKPHGLPPAALHLIAYKENKDRDKRSALS CVISKTARLLSSEDRARLPELRSLVQK 524
Db 1141 LKVIRELOQIAEL-----QEDLES-----EKASRNKAEKQKRDISELEAL--- 1182
Qy 525 RYELLEHKKQWASMSBEORKEYLKKRE-----ELKKKLKEKAK-----ERREK-----568
Db 1183 KTELEDTLDTTAAQOE-----LRTKREQEVALKEKAIBEETKNHEAQIQEIRQRHATAL 1236
Qy 569 -EMLELEKOKRYEDQELGTGNLPAFRIVDTPPEGLPNTLFGDVAMVVEFLSCVGLLLPD 627
Db 1237 BELSEQUEQNGRPF-----ANLEKNQKOGLES-----D 1263
Qy 628 AOYPITAVSLMEALSADKGGFLYNRLVILLQTL-----LODEIAEDYGEIGMKL 678
Db 1264 NKEACEVKVLQVKAESE-----HKRKLDAQVQELTAKYTEGERLVELAE-----KA 1313
Qy 679 SEIPLTLHVSVELVRLCLARSDVOESESDDTDNKSAAAFEDNEVQDEFLEKLETSEFF 738
Db 1314 NKILQNELDNVSSLL-----BEAKKGKIFAKDAASLE-SQLQD-----TOELL 1355
Qy 739 ELTSEKLIQLLTALCHRLMTYSVQDHMETRQMSAELMWERLAVLKEENDKKAQKQ-- 796
Db 1356 QETROKLN-----LSSRIQLEBEKNLQEQOE-----EESBEARKNLEKQML 1398
Qy 797 --KRKEMEAKNGKENGVLGKTRKRVKPEPOVTEADMDISAVKSRLLAIQAKK 854
Db 1399 ALQOALAEAKKVDLDTGTEGLEENKLLKDMESLSQRLSEKANAYDK-----LEKTK 1453
Qy 855 EREIQE-REMKVLEKQAEERIRKHAAAEKAFQEGIAKAKLVMMRTPIGTDNRNRYW 913
Db 1454 NRIQCELDLMDVLDHQ--RQIVSNLEKKKQKFDQMLAEKNISAR--YABERDRAEAE 1508
Qy 914 LFSDEVPGLFIEKGWHDSIDYRFNHCKDHTVSGD-EDYCPRSKKNLGNKASMTQHG 972
Db 1509 AREKETKALSUKA-LEEALEAKEEPERQNKQLRADMEDL--MSSKDDVVGKNVHELEKSK 1565
Qy 973 TATEVAVE---TTTPKQGNLWFLCDSQKELDELLNCLHPQGIRESQ-----1016
Db 1566 RTLQEQVEEMRTQLEEBELQATEDAKLRLVNMQAKMAQFERDLQARDEQNEEKKRML 1625
Qy 1017 -----LKERLEKRYODIITHSILARKPNLGLKCDGNOELNLFRLSDLIEVATRLQKGG 1070
Db 1626 VKQVRELEAELEDERQORALAAVAAKKMEMDLKLEGOIEAANKARDEAIKQLRKLQAM 1685
Qy 1071 LGVVEETSE-----FEARVISLEKLDGFCVIALQASVIKKFLQGFMAPKQKRL 1122
Db 1686 KOYQRELEEARSRDEIFAQSEKSEKKUGLEBAILQLO-----EEFAASERARRHA 1737

Qy 1123 QSE-----DSAKTEYDDEEKKMVEEAKVASALEKWKTAIREAQT-----FS 1163
Db 1738 EQERDELADEIANASAGKSALLDEKRL--EARIAQLEEE--LEEQSNMELLNERFR 1791
Qy 1164 RMHVLLGMLDPAICIKWDMSA-----ENARCKVCPPKGGEDDKLILCDECNKAFHFLCLRPALY 1219
Db 1792 KTTLQVDTLNLSELAGERSAAQKSENARQOL-----ERQNK-----LKAKLQ 1833
Qy 1220 EYPDGEWQCPACOPATARRNSGRNYTESASEDSDDDES-----DEE 1262
Db 1834 EL--EGSVKSKFKATISTLEAKIAQLEEQLEQAEKERAANKLVRTTEKKLKEVPMQVE 1890
Qy 1263 EEEEEEEEDYEVAGLRPRKTIRGHSHVIPPAAASGRPRPGKPKPHSTRRSQPKAPPV 1322
Db 1891 DERRHADQYKEQWENAKANARKOLK--QLEBAEBEATRANASRRKLQR--EL 1938
Qy 1323 DDA-EVDE-----LVLOTGRSSRRQ-----SLEL 1345
Db 1939 DDATANEGLSREVSTLKNLRRGGPITFFSSRSRGRQLHIEGASLEL 1986

Search completed: March 9, 2005, 14:24:56
Job time : 84.0703 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 9, 2005, 13:06:37 ; Search time 121.327 Seconds
(without alignments)
5336.298 Million cell updates/sec

Title: US-10-702-148-1
Perfect score: 8631
Sequence: 1 MEDASRSSRGVAPLINNVL.....VTPSNVDQVSTPPAAKSRI 1674

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues
Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 65 summaries

Database : A_Geneseq_16Dec04:*

1: geneseqp1980s:*

2: geneseqp1990s:*

3: geneseqp2000s:*

4: geneseqp2001s:*

5: geneseqp2002s:*

6: geneseqp2003as:*

7: geneseqp2003bs:*

8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	8631	100.0	1674	2 AAW81169	Aaw81169 Human BAZ
2	8631	100.0	1674	7 ADF69139	Adf69139 Human MP5
3	8631	100.0	1674	8 ADP12578	Adp12578 Protein e
4	7702	89.2	1540	2 AAY07734	Aay07734 Human HTL
5	2437	28.2	475	4 AAB93746	Aab93746 Human pro
6	1280.5	14.8	1476	4 ABB58706	Abb58706 Drosophil
7	1215	14.1	279	5 ABB90253	Abb90253 Human pol
8	945.5	11.0	1483	8 ADQ20910	Adq20910 Human sof
9	938.5	10.9	1483	7 ADF69140	Adf69140 Human MP5
10	930.5	10.8	1531	2 AAW81173	Aaw81173 Human BAZ
11	929.5	10.8	1527	2 AAW81172	Aaw81172 Human BAZ
12	727.5	8.4	1972	2 AAW81171	Aaw81171 Human BAZ
13	726.5	8.4	1972	6 ABR64241	Abf64241 Angiogene
14	726.5	8.4	1972	8 ADP54420	Adp54420 Human PRO
15	693.5	8.0	1589	4 AAM42025	Aam42025 Human pol
16	689.5	8.0	1727	4 AAB95554	Aab95554 Human pro
17	689.5	8.0	1878	4 AAM40239	Aam40239 Human pol
18	688.5	8.0	1873	7 AAM85417	Aam85417 Human pro
19	681.5	7.9	1878	2 AAW81170	Aaw81170 Human BAZ
20	680	7.9	1586	7 AAM85416	Aam85416 Mouse pro
21	662	7.7	149	2 AAY07982	Aay07982 Human sec
22	493	5.7	1052	6 ABR41364	Abf41364 Human DIT
23	447.5	5.2	524	8 ADR09529	Adf09529 Human pro
24	381	4.4	708	4 AAB93073	Aab93073 Human pro
25	368.5	4.3	796	4 AAM93402	Aam93402 Human pol

26	368.5	4.3	796	8 ADL30971	Adl30971 Human PRO
27	361.5	4.2	1398	5 ABG97491	Abg97491 Human NOV
28	361.5	4.2	1400	4 AAB83348	Aab83348 AAP-2 pro
29	361.5	4.2	1445	6 ABR41365	Abf41365 Human DIT
30	343.5	4.0	1584	8 AAM80133	Aam80133 Tumour-as
31	338	3.9	244	4 AAB63263	Abb63263 Human bre
32	326.5	3.8	2246	4 ABG05850	Abg05850 Novel hum
33	318.5	3.7	572	4 AAM14838	Aam14838 Peptide #
34	318.5	3.7	572	4 ABB33805	Abb33805 Peptide #
35	318.5	3.7	572	4 AAM27265	Aam27265 Peptide #
36	318.5	3.7	572	4 ABB28622	Abb28622 Peptide #
37	318.5	3.7	572	4 ABB19248	Abb19248 Protein #
38	318.5	3.7	572	4 AAM66978	Aam66978 Human bon
39	318.5	3.7	572	4 AAM54571	Aam54571 Human bra
40	318.5	3.7	572	4 ABG48640	Abg48640 Human liv
41	318.5	3.7	572	4 AAM02563	Aam02563 Peptide #
42	318.5	3.7	572	5 ABG36635	Abg36635 Human pep
43	317	3.7	513	4 AAM00760	Aam00760 Human bon
44	316	3.7	2274	4 ABB58657	Abb58657 Drosophil
45	314	3.6	560	4 AAM18320	Aam18320 Peptide #
46	314	3.6	560	4 ABB37354	Abb37354 Peptide #
47	314	3.6	560	4 AAM30808	Aam30808 Peptide #
48	314	3.6	560	4 ABB32101	Abb32101 Peptide #
49	314	3.6	560	4 ABB22640	Abb22640 Protein #
50	314	3.6	560	4 AAM70484	Aam70484 Human bon
51	314	3.6	560	4 AAM58044	Aam58044 Human bra
52	314	3.6	560	4 AAM05928	Aam05928 Peptide #
53	314	3.6	560	5 ABG40123	Abg40123 Human pep
54	310	3.6	1898	2 AAY30795	Aay30795 A human t
55	310	3.6	1898	7 ADD48869	Add48869 Human PRO
56	308	3.6	3712	8 ADF45526	Adf45526 Pan trogl
57	306	3.5	2272	2 AAW21731	Aaw21731 GAL4/HA/N
58	306	3.5	3818	8 ADF45525	Adf45525 Human AKA
59	305	3.5	2018	8 ABBM81287	Abbm81287 Tumour-as
60	305	3.5	2117	4 AAU32040	Aau32040 Novel hum
61	305	3.5	2192	2 AAW21732	Aaw21732 LexA/NuMA
62	303.5	3.5	1130	8 ADN24098	Adn24098 Bacterial
63	299	3.5	3257	4 ABB67502	Abb67502 Drosophil
64	297.5	3.4	2897	4 ABB58514	Abb58514 Drosophil
65	297.5	3.4	3907	6 ABR92047	Abf92047 Human cer

ALIGNMENTS

RESULT 1
AAW81169
ID AAW81169 standard; protein; 1674 AA.

XX AAW81169;
AC
DT 05-MAY-1999 (first entry)
XX Human BAZI-alpha protein.

DE Transcriptional regulator; BAZI-alpha; bromodomain; BAZ;
XX atypical zinc finger; testis; human; tumour; BAZI-beta; BAZ2-alpha; drug;
KW BAZ2-beta; treatment; cancer; proliferative disorder; screening.

XX Homo sapiens.
OS
PN WO9847920-A1.

XX 29-OCT-1998.
XX
PF 17-APR-1998; 98WO-JP001783.
XX
PR 18-APR-1997; 97JP-00116570.
PR 24-OCT-1997; 97JP-00310027.

XX (CHUS) CHUGAI RES INST MOLECULAR MEDICINE INC.
XX Jones MH;
PI

7

XX Human MP53 protein sequence SEQ ID NO:109.
 XX p53 pathway modulating agent; MP53; p53 modulator; cytostatic;
 KW gene therapy; cancer; human.
 XX Homo sapiens.
 XX WO2003083047-A2.
 PN 09-OCT-2003.
 XX 28-FEB-2003; 2003WO-US006025.
 XX 01-MAR-2002; 2002US-0361196P.
 XX (EXEL-) EXELIXIS INC.
 PI Belvin M, Francis-Lang H, Friedman L, Plowman GD, Heuer TS, Li D;
 PI Funke RP;
 XX WPI; 2003-812540/76.
 DR N-PSDB; ADF69195.
 XX Identifying a candidate p53 pathway modulating agent for treating e.g.,
 PT cancer by contacting an assay system comprising a MP53 polypeptide or
 PT nucleic acid with a test agent and detecting a test agent-biased
 PT activity.
 XX Example; SEQ ID NO 109; 406pp; English.
 XX The present invention describes a method for identifying a candidate p53
 CC pathway modulating agent, which comprises: (a) providing an assay system
 CC comprising a MP53 (modulator of p53) polypeptide or nucleic acid or its
 CC fragment or derivative; (b) contacting the assay system with a test agent
 CC under conditions where the system provides a reference activity except in
 CC the presence of the test agent; and (c) detecting a test agent-biased
 CC activity, where a difference between the test agent-biased activity and
 CC the reference activity identifies the test agent as a candidate p53
 CC pathway modulating agent. Also described: (1) modulating the p53 pathway
 CC of a cell; (2) modulating the p53 pathway in a mammalian cell; and (3)
 CC diagnosing a disease in a patient. MP53 has cytostatic activity, and can
 CC be used in gene therapy. The method is useful for identifying a candidate
 CC p53 pathway modulating agent for preparing a composition for diagnosing
 CC or treating e.g., cancer. The present sequence represents a human MP53
 CC protein, which is used in the exemplification of the present invention.
 XX Sequence 1674 AA;

Query Match 100.0%; Score 8631; DB 7; Length 1674;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1674; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MEDASESRGVAFLNNVVLPGSPSLPVSVTGCKSHRVANKVVEARSEKLLPTALPPE 60
 Db 1 MEDASESRGVAFLNNVVLPGSPSLPVSVTGCKSHRVANKVVEARSEKLLPTALPPE 60

Qy 61 PKVDQKLPRSSRRSGGGTQFPARSRAVAGEAAAGAPERGSPIGRRVSPRLCSG 120
 Db 61 PKVDQKLPRSSRRSGGGTQFPARSRAVAGEAAAGAPERGSPIGRRVSPRLCSG 120

Qy 121 EGGQVAVGVIAKGRGRGDSRRAPGGREMPLLHRKPFVQKPPADLRPDVEEYFCKVT 180
 Db 121 EGGQVAVGVIAKGRGRGDSRRAPGGREMPLLHRKPFVQKPPADLRPDVEEYFCKVT 180

Qy 181 NEIFRHYDDFFERTILCNLSLVMSCAVTRPGLTYQEALESEKARQNLQSPPELIIPVL 240
 Db 181 NEIFRHYDDFFERTILCNLSLVMSCAVTRPGLTYQEALESEKARQNLQSPPELIIPVL 240

Qy 241 YLTSITHSRLHEICDDIFAYYKDYFVEETVEVIRNNGARLQCTILEVLPSSHONGFAN 300
 Db 241 YLTSITHSRLHEICDDIFAYYKDYFVEETVEVIRNNGARLQCTILEVLPSSHONGFAN 300

Qy 301 GHVNSVDGETIIISDDSETQSCSFQNGKKKDAIDPLLFYKVQVPTKKELHESAIVKAT 360
 Db 301 GHVNSVDGETIIISDDSETQSCSFQNGKKKDAIDPLLFYKVQVPTKKELHESAIVKAT 360

Qy 361 QISRRKGLFSRDKLKLFLKQHCPEQGVIKIKASSLSYKIAEQDPSYFFPDPTTFIFS 420
 Db 361 QISRRKGLFSRDKLKLFLKQHCPEQGVIKIKASSLSYKIAEQDPSYFFPDPTTFIFS 420

Qy 421 PANRRGRPPKRIHISQEDNVANKQTLASYRSKATKRDKLLKQEEMKSLAFKAKLKRE 480
 Db 421 PANRRGRPPKRIHISQEDNVANKQTLASYRSKATKRDKLLKQEEMKSLAFKAKLKRE 480

Qy 481 KADALEAKKKEKDEKKEKREELKIVIEBERLKKKEERLKVVEREKEREKREKRYVE 540
 Db 481 KADALEAKKKEKDEKKEKREELKIVIEBERLKKKEERLKVVEREKEREKREKRYVE 540

Qy 541 YLKQWSKPREMECDDDLKELPEPTPVKTRLPPEIFGDALMWLEFLNAGELFDLQDEPDP 600
 Db 541 YLKQWSKPREMECDDDLKELPEPTPVKTRLPPEIFGDALMWLEFLNAGELFDLQDEPDP 600

Qy 601 GVTLEVEEALVGNDSGGLCELLFFLTAIFQAIABEEBVAKEQLTADTKGCSLKS 660
 Db 601 GVTLEVEEALVGNDSGGLCELLFFLTAIFQAIABEEBVAKEQLTADTKGCSLKS 660

Qy 661 DLDSCTLSLILRLHILASGADVTSANAKRYQKRGGFATDDACMELSLNPSLVKULSS 720
 Db 661 DLDSCTLSLILRLHILASGADVTSANAKRYQKRGGFATDDACMELSLNPSLVKULSS 720

Qy 721 TSVYDLTPGKWKILHALCGKLLTLVSTRDIEDYVILRQAKQEFRELKAEQHKREKE 780
 Db 721 TSVYDLTPGKWKILHALCGKLLTLVSTRDIEDYVILRQAKQEFRELKAEQHKREKE 780

Qy 781 AAAIRKREKLEKEQKMEKQEKLEKEDQKRNSTADISIGEEEREDFDTSIESKDETEQ 840
 Db 781 AAAIRKREKLEKEQKMEKQEKLEKEDQKRNSTADISIGEEEREDFDTSIESKDETEQ 840

Qy 841 KELQDMFTEDDEDPGSHKRRGRRGKQNGKFTQKQINCQVTRRELLTADDEEALKQEH 900
 Db 841 KELQDMFTEDDEDPGSHKRRGRRGKQNGKFTQKQINCQVTRRELLTADDEEALKQEH 900

Qy 901 QRKEKLEKIQSAIACTNIEPLGRDRMYRYWIFPSIPGLFIEDYSGLTEDMLPSPS 960
 Db 901 QRKEKLEKIQSAIACTNIEPLGRDRMYRYWIFPSIPGLFIEDYSGLTEDMLPSPS 960

Qy 961 SFQNNVQSDPQVSTKTGCEPLMSESTNIDQGRDHSVQLPKPVHKPNRMCFYSCEQLD 1020
 Db 961 SFQNNVQSDPQVSTKTGCEPLMSESTNIDQGRDHSVQLPKPVHKPNRMCFYSCEQLD 1020

Qy 1021 QLIBALNSRGHRESALKETLLQEKSRICAQLARFSEKHFSDKPPQDSKPTYSRGRSSN 1080
 Db 1021 QLIBALNSRGHRESALKETLLQEKSRICAQLARFSEKHFSDKPPQDSKPTYSRGRSSN 1080

Qy 1081 AYDPSQCAEKQELRLRDLFDIEDRIYQGTGLGAIKVTDRHWRSALESGRYELLSEEN 1140
 Db 1081 AYDPSQCAEKQELRLRDLFDIEDRIYQGTGLGAIKVTDRHWRSALESGRYELLSEEN 1140

Qy 1141 KENGIITKTVNEDVEMEIDEQTKIVKORLLGKITETPSTVSTNASTPQSVSSVHYLAM 1200
 Db 1141 KENGIITKTVNEDVEMEIDEQTKIVKORLLGKITETPSTVSTNASTPQSVSSVHYLAM 1200

Qy 1201 ALFQIEQGIERRFLKAPLDASDSGRSYKTVLDRWRRESLLSSASISQVFLHLSTLDRSVI 1260
 Db 1201 ALFQIEQGIERRFLKAPLDASDSGRSYKTVLDRWRRESLLSSASISQVFLHLSTLDRSVI 1260

Qy 1261 SKSILNARCKIKRKGDANNVLCGGDRGHHTVCVRPKLKTVPEDMFCPECRCQRCR 1320
 Db 1261 SKSILNARCKIKRKGDANNVLCGGDRGHHTVCVRPKLKTVPEDMFCPECRCQRCR 1320

Qy 1321 RLSFRQPSLESDEVDSDMGDEVDGDEEQSEEEVEVEQDEDDSDQEEBVSUPK 1380
 Db 1321 RLSFRQPSLESDEVDSDMGDEVDGDEEQSEEEVEVEQDEDDSDQEEBVSUPK 1380

Qy 1381 RGRPQVRLPVKTRGKLSSSFSSRQOQBPGRYPSPRSQOSTPKTTVSSKTGRSLRINSAP 1440

Db 1381 RGRPOVRLPVKTRGLSSFSRGGQQEPGRYPSSQQSTPKTTVSSKTRSLRKINSAP 1440
Qy 1441 PTEKSLRIASRSTRSHSGPLOADVVELLSPPRRKRGRKSNANTPNSPNPFRVIAT 1500
Db 1441 PTEKSLRIASRSTRSHSGPLOADVVELLSPPRRKRGRKSNANTPNSPNPFRVIAT 1500
Qy 1501 KSSEQSRVNTASKLSLOESKRRCKRQSPSPVTLGRRSSGRQGVHLSAFEOLV 1560
Db 1501 KSSEQSRVNTASKLSLOESKRRCKRQSPSPVTLGRRSSGRQGVHLSAFEOLV 1560
Qy 1561 VELVRHDSWPFLKLVSKIQVPDYDIIKKPIALNIIREKNKCEYKLASEFIDDIELMF 1620
Db 1561 VELVRHDSWPFLKLVSKIQVPDYDIIKKPIALNIIREKNKCEYKLASEFIDDIELMF 1620
Qy 1621 SNCFEYNPRNTSEAKAGTRLOAFPHIOAKLGLHVTSPNVQVSTPPAAKKSRI 1674
Db 1621 SNCFEYNPRNTSEAKAGTRLOAFPHIOAKLGLHVTSPNVQVSTPPAAKKSRI 1674

RESULT 3

ID ADP12578 standard; protein; 1674 AA.
XX AC ADP12578;
XX 12-AUG-2004 (first entry)
XX DE Protein encoded by mRNA of the invention #188.
XX transplamt rejection; immune system; rheumatoid arthritis; lupus;
KW inflammatory bowel disease; multiple sclerosis; HIV; AIDS.
XX Homo sapiens.
XX WO2004042346-A2.
XX 21-MAY-2004.
XX 24-APR-2003; 2003WO-US012946.
XX 24-APR-2002; 2002US-00131831.
XX 20-DEC-2002; 2002US-00325899.
XX (EXPR-) EXPRESSION DIAGNOSTICS INC.
XX Wohlgenuth J, Fry K, Woodward R, Ly N, Prentice J, Morris M;
PI Rosenberg S;
XX WPI; 2004-400724/37.

DR WPI; 2004-400724/37.
XX Diagnosing or monitoring transplant rejection, e.g. heart, kidney, liver,
PI pancreas, pancreatic islet, lung, bone marrow or stem cell transplant
PI rejection, in an individual, comprises detecting the expression level of
XX the genes.
XX Claim 65; SEQ ID NO 2587; 1762pp; English.
XX The present invention relates to diagnosing or monitoring transplant
XX rejection, e.g. cardiac or kidney transplant rejection, in an individual
XX comprises detecting the expression level of one or more genes. The
XX methods, system and kits are useful in diagnosing or monitoring
XX transplant rejection, e.g. heart, kidney, liver, pancreas, pancreatic
XX islet, lung, bone marrow or stem cell transplant rejection,
XX xenotransplant rejection or mechanical organ replacement rejection, in an
XX individual. The method is also useful in assessing the immune status of
XX an individual. The methods are also useful in diagnosing and monitoring
XX diseases that involve the immune system, e.g. rheumatoid arthritis,
XX lupus, inflammatory bowel diseases, multiple sclerosis, HIV/AIDS or
XX viral, bacterial or fungal infection. The present sequence represents a
XX protein that is encoded by the mRNA of the invention.
XX Sequence 1674 AA;

Query Match		100.0%;	Score 8631;	DB 8;	Length 1674;
Best Local Similarity		100.0%;	Pred. No. 0;		
Matches 1674;		Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	MEDASESSRGVAPLINNVNVLPGSPSLPVSVTGCKSHRVANKKVEARSEKLLPTALPPSE	60		
Db	1	MEDASESSRGVAPLINNVNVLPGSPSLPVSVTGCKSHRVANKKVEARSEKLLPTALPPSE	60		
Qy	61	PKVDQKLPRESRRGGSGGTQFPARSRAVAAGEAAAAAGAPRGSPIGRRVSPRCLCSG	120		
Db	61	PKVDQKLPRESRRGGSGGTQFPARSRAVAAGEAAAAAGAPRGSPIGRRVSPRCLCSG	120		
Qy	121	EGGOVAVGTAGKGRGDRGSRAPCGREWPLLRKPFVROKPPADLRDDEEVFYCKVT	180		
Db	121	EGGOVAVGTAGKGRGDRGSRAPCGREWPLLRKPFVROKPPADLRDDEEVFYCKVT	180		
Qy	181	NEIFRHYDDFFERTILCNSLVMSCAVTGRPGLTVQEALESEKKARONLQSPPEPLIIPVL	240		
Db	181	NEIFRHYDDFFERTILCNSLVMSCAVTGRPGLTVQEALESEKKARONLQSPPEPLIIPVL	240		
Qy	241	YLTSLTHRSRLHEICDDIFAYVKDRYFVEETVEVRNNGARLOCTILEVLPPSHONGFAN	300		
Db	241	YLTSLTHRSRLHEICDDIFAYVKDRYFVEETVEVRNNGARLOCTILEVLPPSHONGFAN	300		
Qy	301	GHVNSVDGETIIISDSDSETQSCSFONGKKDAIDPLLFYKVQPTKKELHESAIYKAT	360		
Db	301	GHVNSVDGETIIISDSDSETQSCSFONGKKDAIDPLLFYKVQPTKKELHESAIYKAT	360		
Qy	361	QISRRKHLFSRDKLFLKQHCPEQEGVIKASSISTYKIAEQDFSYFFDDPTTIFS	420		
Db	361	QISRRKHLFSRDKLFLKQHCPEQEGVIKASSISTYKIAEQDFSYFFDDPTTIFS	420		
Qy	421	PANRRGRPPKRIHISOEDNVANKQTLASVRSKATKERDKLLKQEEWKSLAFKAKLKRE	480		
Db	421	PANRRGRPPKRIHISOEDNVANKQTLASVRSKATKERDKLLKQEEWKSLAFKAKLKRE	480		
Qy	481	KADALEAKKKEKEDKEKKREELKKIVEERLKKYKEERLKVREKEREKLEERKKYVE	540		
Db	481	KADALEAKKKEKEDKEKKREELKKIVEERLKKYKEERLKVREKEREKLEERKKYVE	540		
Qy	541	YLKQWSKPREDMECDDLKELPEPTPVKTRLPPEIFGDALMVLEFLNAGLFLDQDFPD	600		
Db	541	YLKQWSKPREDMECDDLKELPEPTPVKTRLPPEIFGDALMVLEFLNAGLFLDQDFPD	600		
Qy	601	GVTLEVLREALVGNDSGPLCELLFFPLTAIFAQIAEBEEVEVAKEQLTDATDKGCSLKSL	660		
Db	601	GVTLEVLREALVGNDSGPLCELLFFPLTAIFAQIAEBEEVEVAKEQLTDATDKGCSLKSL	660		
Qy	661	DLDSCTLSEILRLHLILASGADVTSANAKYRYOKRGGFDTDDACMEILNSPVLKSLSS	720		
Db	661	DLDSCTLSEILRLHLILASGADVTSANAKYRYOKRGGFDTDDACMEILNSPVLKSLSS	720		
Qy	721	TSVYDLTPGKMKILHALCGKLLTLVSTRDFIEDYVDILRQAKQEFRELKAEQHRKERE	780		
Db	721	TSVYDLTPGKMKILHALCGKLLTLVSTRDFIEDYVDILRQAKQEFRELKAEQHRKERE	780		
Qy	781	AAAIRRKEEKLKEQEQKMKQEKLKEDEQRNSTADISIGBEEREDFDTIESKDTEQ	840		
Db	781	AAAIRRKEEKLKEQEQKMKQEKLKEDEQRNSTADISIGBEEREDFDTIESKDTEQ	840		
Qy	841	KELDQDMFTDEDDPGSHKGRGKRGQNGFKFTQEQINCVTRELLTADAEALQEH	900		
Db	841	KELDQDMFTDEDDPGSHKGRGKRGQNGFKFTQEQINCVTRELLTADAEALQEH	900		
Qy	901	QRKEKELLEKIQSAIACTNIFPLGRDMYRYWIFPSIPGLFIEEDYSGLTDMLLPRPS	960		
Db	901	QRKEKELLEKIQSAIACTNIFPLGRDMYRYWIFPSIPGLFIEEDYSGLTDMLLPRPS	960		
Qy	961	SFQNVQSQDPQVSTKTGTEPLMSESTSNIDQGRDHSVQLPKPVHKPNRRCFYSSCSOLD	1020		
Db	961	SFQNVQSQDPQVSTKTGTEPLMSESTSNIDQGRDHSVQLPKPVHKPNRRCFYSSCSOLD	1020		

QY 1021 QLIETALNSRGHRESALKETTLQEKSRICAOIARFSEKHFSDKPODPKPTYSKGRSSN 1080
 Db 1021 QUIETALNSRGHRESALKETTLQEKSRICAOIARFSEKHFSDKPODPKPTYSKGRSSN 1080
 QY 1081 AYDPSQMAEKQLELRDLFDLIDEDRIYQGTGLGAKVTDRIHWRSALESGRYELLSEN 1140
 Db 1081 AYDPSQMAEKQLELRDLFDLIDEDRIYQGTGLGAKVTDRIHWRSALESGRYELLSEN 1140
 QY 1141 KENGIKTVNEDVEEMIDEQKVIVKORLLGIKTETPTSTVSTNASTPQSVSSVHYLAM 1200
 Db 1141 KENGIKTVNEDVEEMIDEQKVIVKORLLGIKTETPTSTVSTNASTPQSVSSVHYLAM 1200
 QY 1201 ALFQIEQGLERFLKAPDASDGRSYKTVLDWRRESLLSSASLOVFLHLSLIDRSVIW 1260
 Db 1201 ALFQIEQGLERFLKAPDASDGRSYKTVLDWRRESLLSSASLOVFLHLSLIDRSVIW 1260
 QY 1261 SKSILNARCKICRKKGDAENMVLCDCDRGHHTYCVRPKLKTVPBGDMFCPECRPKQCR 1320
 Db 1261 SKSILNARCKICRKKGDAENMVLCDCDRGHHTYCVRPKLKTVPBGDMFCPECRPKQCR 1320
 QY 1321 RLSFRQPSLESDEVDVDSMGEDDEVDCDEEGQSEEEYEVEQDEDDSOEEEVSLPK 1380
 Db 1321 RLSFRQPSLESDEVDVDSMGEDDEVDCDEEGQSEEEYEVEQDEDDSOEEEVSLPK 1380
 QY 1381 RGRPOVRLPVKTRGKLSSFSRSGQQOEPGRYPSPRSQOSTPKTTVSSKTGRSLRKINSAP 1440
 Db 1381 RGRPOVRLPVKTRGKLSSFSRSGQQOEPGRYPSPRSQOSTPKTTVSSKTGRSLRKINSAP 1440
 QY 1441 PTEKSLRIASRSTRHSHGPIQADVVELLSPRRKRGRKSGANNTPENS PNFPNFRVIAT 1500
 Db 1441 PTEKSLRIASRSTRHSHGPIQADVVELLSPRRKRGRKSGANNTPENS PNFPNFRVIAT 1500
 QY 1501 KSSEQSRVNIASKLSLOESKRRCKRQSPSPVTLGRRSSQGGVHLSAFEOLV 1560
 Db 1501 KSSEQSRVNIASKLSLOESKRRCKRQSPSPVTLGRRSSQGGVHLSAFEOLV 1560
 QY 1561 VELVRHDSWPFLKLVSKIQVDDYDIKKPIALNIIREKVNKCEYKLASEFIDDIELMF 1620
 Db 1561 VELVRHDSWPFLKLVSKIQVDDYDIKKPIALNIIREKVNKCEYKLASEFIDDIELMF 1620
 QY 1621 SNCFEYNPNRTSEAKAGTRLOAFHHIOAKLGLHVTSPNVQVSTPPAAKKSRI 1674
 Db 1621 SNCFEYNPNRTSEAKAGTRLOAFHHIOAKLGLHVTSPNVQVSTPPAAKKSRI 1674

RESULT 4

AAV07734
 ID AAV07734 standard; protein; 1540 AA.

AC AAV07734;

DT 02-JUL-1999 (first entry)

DE Human hTLP2 protein.

KW Telomerase protein; hTLP2; human; biocontrolling mechanism; cell growth;
 cell aging.

OS Homo sapiens.

PN JP11089579-A.

PD 06-APR-1999.

PF 24-SEP-1997; 97JP-00258450.

PR 24-SEP-1997; 97JP-00258450.

PA (MITU) MITSUBISHI CHEM CORP.

XX WPI; 1999-290828/25.

DR N-FSDB; AAX37304.

XX

PT A higher animal telomerase protein and a gene coding it.

XX Example 1; Page 12-18; 22pp; Japanese.

CC This invention describes a novel human telomerase protein, hTLP2. This
 protein and the gene coding it are useful for the elucidation of
 biocontrolling mechanisms such as cell growth and cell aging

XX Sequence 1540 AA;

Query Match 89.2%; Score 7702; DB 2; Length 1540;

Best Local Similarity 96.5%; Pred. No. 0;

Matches 1502; Conservative 1; Mismatches 5; Indels 48; Gaps 2;

QY 151 MPLLHRPFVQKPPADLRPDDEEVYKVTNEIFRHYDDDFERTILCNLSVMSCAVTCRP 210

Db 1 MPLLHRPFVQKPPADLRPDDEEVYKVTNEIFRHYDDDFERTILCNLSVMSCAVTCRP 60

QY 211 GLTYQEALESSEKARQNLQSPPELIIIPVLTSLTHRSRLHEICDDIFATVVKDYFYVEE 270

Db 61 GLTYQEALESSEKARQNLQSPPELIIIPVLTSLTHRSRLHEICDDIFATVVKDYFYVEE 120

QY 271 TVEVIRNNGARLQCTILEVLPPSHQNGFANGHVNSVDGETIIISDSDSETQSCSFQNGK 330

Db 121 TVEVIRNNGARLQCTILEVLPPSHQNGFANGHVNSVDGETIIISDSDSETQSCSFQNGK 180

QY 331 KKDAIDPLLFYKQVPTKKELHESAIVKATQISRRKHLFSRDKLKLFLKQHCPEQGVIK 390

Db 181 KKDAIDPLLFYKQVPTKKELHESAIVKATQISRRKHLFSRDKLKLFLKQHCPEQGVIK 240

QY 391 IKASSLSTYKIAEODFSYFPDPPPTIFSPANRRRGRPPKRIHISOEDNVANKQTLAS 450

Db 241 IKASSLSTYKIAEODFSYFPDPPPTIFSPANRRRGRPPKRIHISOEDNVANKQTLAS 300

QY 451 RSKATKRDKLLKQEMKSLAFKAKLREKADALEAKKKEKEDKREKREKLVVEER 510

Db 301 RSKATKRDKLLKQEMKSLAFKAKLREKADALEAKKKEKEDKREKREKLVVEER 360

QY 511 LKKEEKERLKVREKEREKREKRYVEYLKQWSPREDMECDLKLPEPTPVKTRL 570

Db 361 LKKEEKERLKVREKEREKREKRYVEYLKQWSPREDMECDLKLPEPTPVKTRL 420

QY 571 PPEIFGDALMVLEFLNAGELFDLQDEFPDQVTVLEALVGNDSQGLCELLFFFLTA 630

Db 421 PPEIFGDALMVLEFLNAGELFDLQDEFPDQVTVLEALVGNDSQGLCELLFFFLTA 480

QY 631 IFQAIABEEBEVAKEQLTDADTK-----GCSLK 658

Db 481 IFQAIABEEBEVAKEQLTDADTK-----GCSLK 540

QY 659 SLDDSDCTSLSEILRLHILASGADVTSANAKYKRGGFQATDDACMELRLSNPSLVKKL 718

Db 541 SLDDSDCTSLSEILRLHILASGADVTSANAKYKRGGFQATDDACMELRLSNPSLVKKL 600

QY 719 SSTSVYDLTGEKWKILHALCGKLLTLVSTRDFTYVDILRQAKQFRELKASQHRKER 778

Db 601 SSTSVYDLTGEKWKILHALCGKLLTLVSTRDFTYVDILRQAKQFRELKASQHRKER 660

QY 779 EEAARIRKKEEKLKEQEQMKKEQKLEDEORNSTADISIGEEREDPDTIESKDT 838

Db 661 EEAARIRKKEEKLKEQEQMKKEQKLEDEORNSTADISIGEEREDPDTIESKDT 720

QY 839 EQKELDDQMTFEDDDPGSHKRGRRGRKGNGKGFKEFTRQEQINCQVTRRLTADBEALKQ 898

Db 721 EQKELDDQMTFEDDDPGSHKRGRRGRKGNGKGFKEFTRQEQINCQVTRRLTADBEALKQ 780

QY 899 EHQRKEKELLEKIQSAIACTNIIPLGRDRMYRRYWIIFPSIPLGFIEDYSGLTEDMLLPR 958

Db 781 EHQRKEKELLEKIQSAIACTNIIPLGRDRMYRRYWIIFPSIPLGFIEDYSGLTEDMLLPR 840

QY 959 PSSFQNNVQSDQPOVSTKTGEPINSESTSNIDQPRDHSVOLPKPVKPNRWCFYSCEQ 1018

Db 841 PSSFQNNVQSDQPOVSTKTGEPINSESTSNIDQPRDHSVOLPKPVKPNRWCFYSCEQ 900

```
QY 1019 LDQIEALNSGRHSALKETLLQEKSRICQAQARFSEKPHFSDKPODPDSKPTYSRGRS 1078
DB 901 LDQIEALNSGRHSALKETLLQEKSRICQAQARFSEKPHFSDKPODPDSKPTYSRGRS 960
QY 1079 SNAYDPSQWCAEKQLELRDLFLDIEDRIYQGTGAIKVTDRIWRSALSGRYELLSE 1138
DB 961 SNAYDPSQWCAEKQLELRDLFLDIEDRIYQGTGAIKVTDRIWRSALSGRYELLSE 1020
QY 1139 ENKNGIITKYNEDVEENEIDBOTKVIIVKDRLLGIKHTPTSTVSTNASTPOSVSVVHYL 1198
DB 1021 ENKNGIITKYNEDVEENEIDBOTKVIIVKDRLLGIKHTPTSTVSTNASTPOSVSVVHYL 1080
QY 1199 AMALFOIEQGIERRFLKAPLDASDSGRYKTVLDWRRESLLSSASLSQVFLHLSTLDRSV 1258
DB 1081 AMALFOIEQGIERRFLKAPLDASDSGRYKTVLDWRRESLLSSASLSQVFLHLSTLDRSV 1140
QY 1259 IWSKSILNARCKICRKGDAENMVLCDGCDRGHHTYCVRPKLTVPEDGDFCPCRPKQR 1318
DB 1141 IWSKSILNARCKICRKGDAENMVLCDGCDRGHHTYCVRPKLTVPEDGDFCPCRPKQR 1200
QY 1319 CRRLSFRORPSLEDEDVDSMGDEDDVDGDDEEGQSEEEYEVEQDEDDSDSEEEVSL 1378
DB 1201 SRRLSRQRPSLEDEDVDSMGDEDDVDGDDEEGQSEEEYEVEQDEDDSDSEEEVSL 1252
QY 1379 PKRGRPQVRLPVKTRGKLSFSSRGQQQEPGRYPSPRSQQSTPKTTVSSKTGRSLRKINS 1438
DB 1253 -----RLPVKTRGKLSFSSRGQQQEPGRYPSPRSQQSTPKTTVSSKTGRSLRKINS 1304
QY 1439 APPTETKSLRIASRSTRHSHGFLQADVVELLSPPRRKRGRKSANNTPENSPNPNFRVI 1498
DB 1305 APPTETKSLRIASRSTRHSHGFLQADVVELLSPPRRKRGRKSANNTPENSPNPNFRVI 1364
QY 1499 ATKSEQSRSNVNIASKLSQSESKRRCRKQSPSPVTLGRRSSRGQGVHELSAFEQ 1558
DB 1365 ATKSEQSRSNVNIASKLSQSESKRRCRKQSPSPVTLGRRSSRGQGVHELSAFEQ 1424
QY 1559 LVVELVRHDDSWPFLKLSKIQVPDYDIKKPTALNIIREKVNKCEYKLASEFIDDIEL 1618
DB 1425 LVVELVRHDDSWPFLKLSKIQVPDYDIKKPTALNIIREKVNKCEYKLASEFIDDIEL 1484
QY 1619 MFSNCFEYNPRNTSEAKAGTRLOAFHHIQAOKLGLHVTSPSNVDQVSTPPAAKKSRI 1674
DB 1485 MFSNCFEYNPRNTSEAKAGTRLOAFHHIQAOKLGLHVTSPSNVDQVSTPPAAKKSRI 1540

RESULT 5
AAB93746
ID AAB93746 standard; protein; 475 AA.
AC AAB93746;
XX
XX
XX
DT 26-JUN-2001 (first entry)
XX
DE Human protein sequence SEQ ID NO:13405.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
XX
XX 07-FEB-2001.
XX
XX 28-JUL-2000; 2000EP-00116126.
XX
XX 29-JUL-1999; 99JP-00248036.
PR 27-AUG-1989; 98TP-00300253.
PR 11-JAN-2000; 2000JP-00118776.
PR 02-MAY-2000; 2000JP-00183767.
PR 09-JUN-2000; 2000JP-00241899.
XX
XX (HELI-) HELIX RES INST.
```

```
XX Ota T. Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Iehi S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI; 2001-318749/34.
DR
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602 full-
PT length cDNAs defined in the specification, and for the detection and/or
PT diagnosis of the abnormality of the proteins encoded by the full-length
PT cDNAs.
XX
XX Claim 8; SEQ ID NO 13405; 2537pp + Sequence Listing; English.
PS
XX
XX The present invention describes primer sets for synthesizing 5602 full-
CC length cDNAs defined in the specification. Where a primer set comprises:
CC (a) an oligo-dr primer and an oligonucleotide complementary to the
CC complementary strand of a polynucleotide which comprises one of the 5602
CC nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in the
CC specification. The primer sets can be used in antisense therapy and in
CC gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893
CC represent human amino acid sequences; and AAH13629 to AAH13632 represent
CC polynucleotides, all of which are used in the exemplification of the
CC present invention
XX
XX SQ Sequence 475 AA;
XX
XX Query Match 28.2%; Score 2437; DB 4; Length 475;
XX Best Local Similarity 99.4%; Pred. No. 1.2e-160;
XX Matches 472; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1200 MALFOIEQGIERRFLKAPLDASDSGRYKTVLDWRRESLLSSASLSQVFLHLSTLDRSVI 1259
DB 1 MALFOIEQGIERRFLKAPLDASDSGRYKTVLDWRRESLLSSASLSQVFLHLSTLDRSVI 60
QY 1260 WSKSILNARCKICRKGDAENMVLCDGCDRGHHTYCVRPKLTVPEDGDFCPCRPKQR 1319
DB 61 WSKSILNARCKICRKGDAENMVLCDGCDRGHHTYCVRPKLTVPEDGDFCPCRPKQR 120
QY 1320 RRLSFRORPSLEDEDVDSMGDEDDVDGDDEEGQSEEEYEVEQDEDDSDSEEEVSLP 1379
DB 121 RRLSFRORPSLEDEDVDSMGDEDDVDGDDEEGQSEEEYEVEQDEDDSDSEEEVSLP 180
QY 1380 KRGRPQVRLPVKTRGKLSFSSRGQQQEPGRYPSPRSQQSTPKTTVSSKTGRSLRKINS 1439
DB 181 KRGRPQVRLPVKTRGKLSFSSRGQQQEPGRYPSPRSQQSTPKTTVSSKTGRSLRKINS 240
QY 1440 PPTETKSLRIASRSTRHSHGFLQADVVELLSPPRRKRGRKSANNTPENSPNPNFRVIA 1499
DB 241 PPTETKSLRIASRSTRHSHGFLQADVVELLSPPRRKRGRKSANNTPENSPNPNFRVIA 300
QY 1500 TKSEQSRSNVNIASKLSQSESKRRCRKQSPSPVTLGRRSSRGQGVHELSAFEQ 1559
DB 301 AKSEQSRSNVNIASKLSQSESKRRCRKQSPSPVTLGRRSSRGQGVHELSAFEQ 360
QY 1560 VVELVRHDDSWPFLKLSKIQVPDYDIKKPTALNIIREKVNKCEYKLASEFIDDIELM 1619
DB 361 VVELVRHDDSWPFLKLSKIQVPDYDIKKPTALNIIREKVNKCEYKLASEFIDDIELM 420
QY 1620 FSNCFEYNPRNTSEAKAGTRLOAFHHIQAOKLGLHVTSPSNVDQVSTPPAAKKSRI 1674
DB 421 FSNCFEYNPRNTSEAKAGTRLOAFHHIQAOKLGLHVTSPSNVDQVSTPPAAKKSRI 475
```


RESULT 6

ABB58706
ID ABB58706 standard; protein; 1476 AA.

XX ABB58706;

XX 26-MAR-2002 (first entry)

XX Drosophila melanogaster polypeptide SEQ ID NO 2910.

XX Drosophila; developmental biology; cell signalling; insecticide; pharmaceutical.

XX Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US009231.

XX 23-MAR-2000; 2000US-0191637P.

XX 11-JUL-2000; 2000US-00614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

XX N-PSDB; ABL02809.

XX New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signaling and cell-cell interactions.

XX Disclosure; SEQ ID NO 2910; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-ABB72072). The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX Sequence 1476 AA;

Query Match 14.8%; Score 1280.5; DB 4; Length 1476;

Best Local Similarity 24.4%; Pred. No. 2.2e-79; Mismatches 616; Indels 367; Gaps 46; Matches 408; Conservative 283;

QY 151 MPELLRKPP-VRQKPPAD--LRPDEVFYCKVTNIFRHYDDFFERTILCNLSVMSCAVT 207

DB 1 MPICKREGFDLNQKSGKNETFDNDQVFCYITKIFRDEYHYFRHVMVINSTVWQCEAT 60

QY 208 GRPGITYQALSEKKAQNLSQFPPELIIPVLYTSLTHRSRLHEICDDIPAYVKDYRF 267

DB 61 GKENLTYEAVKSERAAKMEQFKQSLRAPVLLVVEHAQQSAVNTLNNIVAKFLRKRYF 120

QY 268 VEETVEVIRNNGARLQCTILEVLPSSHQNGFANGHNSVDGTTIISDSDSETSCSFQ 327

DB 121 IGEESVQAKKNATY--TVLGVKLDKNMPEPLNG-----IYEDTN----- 159

QY 328 NGKKKDAIDPLLFYKVKQPTKKELHESAIKVKATQISRRKHLPSRDLKFLKQHCPEQBG 387

DB 160 -----LVRLRNKGDPSAELDLPRQLRRRMMENLENLWFTKSNVSRVDG 207

QY 388 VIKIKASSLSTYKIAEQ--DFSYPFPDPPPTFFSPANRRGRPPKRIHISOEDNVANKQ 445

DB 208 LLRPKEAYQY-VTDPGVNFSTTIFIGKQPR--YSPAKIKKPDGKKQ-----STLNKY 257

QY 446 TLASYRSKATKERDKLLQBEEMKSLAFKAKADALEAKKKEKEDKEKKREBELKKI 505

DB 258 1VA---GEATAAKSAKAKSDAKSLAELEERVKREK-----EAKLIELE----- 298

QY 506 VEEERLKKKEERLKVREKERERKLEEKKKYVEYLKQMSKPRDEMCDDLKELPEPTP 565

DB 299 -----KQKAEEKQALIERVENECNLLQK-----TDDLERTDQKVLPRYRQ 339

QY 566 VKTRLPEIFGDALMVLEFLNAFGELFDLQDEFPDGVTLVELEALVGNDSGELCELLF 625

DB 340 IVTLPEHLGDAFMREFMHTYTGLSGIEVFQRNLSFYEMTRALTAREIAGPLUSDILL 399

QY 626 FFLTAIFQAIABEEREE-----VAKEQLTDADTKGCSLSLDLDS 664

DB 400 VLLGTVPDLQKEEEREECAVYLDRAAQOEYWSMAQAKSHLYAKRIFSPKVNELPDA 459

QY 665 CTLSEILRLHLIAGADVTSANAKYQKRGFGFATDDACMELRNSPLSVKLSSTSVY 724

DB 460 LTLSEVLRHLHLLGSGAFVNEKAERWVRMYRNGYSKEDPGLELRLEHSHILRLKNHVS 519

QY 725 DLTPEKMKILHALCGKLLTLVSTDFIEDYVDILRQAKQEFRELKAEQHRKEREAAAR 784

DB 520 QLKFKDINLLIRCLMSQIMTYSGTINLIEERMEBOTAKARQDLRALVVGENKR---LAAVE 576

QY 785 IRKKEKEKKEQE-----QMKKEQEKLEDEQRNSTADISIGEEREDFDTSESOT 838

DB 577 INRKLITQMHLEVNGVEPEKREALVEKLLK-----SIAB----- 611

QY 839 EQKELDQDMFTEDDDPGSHKRRGRKRGONGKPFKEFTRQEQINCVTRELLTADBEALQ 898

DB 612 -----LHAQSD-----Q 618

QY 899 EHQRKEKELLEKIOSAICTNIPFLGRDRMYRRWIFPSIPLFTEEDYSGLTEDMLPR 958

DB 619 QHRKHELQML-KLHSQLFNFVLVY-LGMDCRYKYVLESMPGIFVEHSPDSL--DTCLEQ 674

QY 959 PSSFQNNVQSODPQVSTKTGPELM-----SESTSNIDQ 991

DB 675 PITNKSQIEIRQSQALPKNRKDLVLLKLYGDDEKTKKAKKHSLENKENQEHLNCSA 734

QY 992 GPRDHSVOLPK-PVH-----KPNRWCFYSSCSQDLQLEALNSR 1029

DB 735 EPMVDVESDPEAPTHFELLMCSDGKRSICVHDSRNGQRQWAYIYKABEIDELIKALPN 794

QY 1030 CHRRESALKE-----TLLOEKSRIC-AQLARFSEKHFSDKPPQDPSKPTYSRGRSS--- 1079

DB 795 GLREYELLQELSVLRSLEHQHAKTCPVDLLSLENETWRKPFMAAMESETNRKYGEANFGL 854

QY 1080 -NAYDPSQMAEKOLELRDRDFLDIEDRIYQGTILGAIKVTDRHWRSALSGRYE--- 1134

DB 855 PNGTDLNEV-----MRLHLVDRIIQFENDIYTGDLGRLLKVKDKMEKWRSDLLGGNYDAOCK 909

QY 1135 -----LLSEENKENGIIKTVNEDVEEMEIDQTKVIVKD--RLGIKTETFTSTVN 1184

DB 910 LQWPGGKLEDEAGSDNESHETHEED-DGALLGVARKPYRDPGMYLAASADTKPLPDS 968

QY 1185 ASTQSVSSV-----VHYLAMALFOIEQIERFLKAPLD-----ASDSGRS 1226

DB 969 DEEDQHTNAVAPIAVHNMASSALLQVEQAIKGRFLKEPYGMKGWDPKQOALKLACDSR-- 1026

QY 1227 YKTVLDWRRESLLSASISQVFLHSLTDRSVWSKSLNARCKTCRKGGAENNVLCDG 1286

DB 1027 ----LHOMVESLMESTSFAQVFLHNLHLDICQWRNSTKSLCKVCRGSDPEKMLLCDE 1082

QY 1287 CDRGHHTYCVRPKLTVPEDGWFCEPC-----RPQRCRLRSFQRPSLESDVE 1337

DB 1083 CNAGTHMFLKPLKRSVPPGNWYCNDCVKSLGSLGQNEKDKKQATKKRKFIVEED-- 1140

QY 1338 DSMGGEDEVGDDEEGQ-----SEEBEYVEQEDDSDQEEBEV-SLPKRGPRQVRLPVK 1391

Db 1141 -----DEATDEEBEKKDDMTDDEDAEHENEKHHDEVEDDESVTSTPSSSRVNGRIILRR 1194
Qy 1392 TRGLSSFSRSGOQOQGRYPSPRSQQSTPKTTVSSKTRGSLRKINSAPPTETSLRIAS 1451
Db 1195 PRTRTSRLRLTSKEIEE-----HAQEDVSDGVSDDA--SLTAGEDTIEDESEKVCQ 1246
Qy 1452 R-----STRSHSGPLQADVVELLS--PRRKRGRKSGANTTPNSP 1490
Db 1247 KCFYDGGIHKVQCRLFFHLECVHLKPRPRIDFVCKTKCPMPQPRRHSNMGDHRDE 1306
Qy 1491 NFPNFRVIATKSSQSRSVNTASKLSQESKSKRKRQSPSPSVTLGRSSRGQGV 1550
Db 1307 BEP-----KAKRPNRSLRLSIDKTARPSNGNNNNNNNS--SVNNNNHRRSGRTNE 1356
Qy 1551 H---ELSAFEQLVVELVRHDSWPLKLVSKIQVDDYDIIKKPTANIIIEKYNKCEYK 1607
Db 1357 HPLNSAALYDLLEQIMGHKAAMPFLPVLTVSEPDYHOIITKPMDLAKIKSLNMGAYQ 1416
Qy 1608 LASEFIDIELMFSNCFEYNPRNTSEAKAGTRLOAFHIOAKGLGHLVTPSNVD 1661
Db 1417 LNEELLSDIQLVFRNCDIYVNEGNEIYDAGCQLERFVIDRCDMQLPFRPSDMN 1470

RESULT 7
ID ABB90253 standard; protein; 279 AA.
XX ABB90253;
AC ABB90253;
XX
DT 24-MAY-2002 (first entry)
XX
DE Human polypeptide SEQ ID NO 2629.
XX
KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;
KW vulnary; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW neurological disease; infection; human; secreted protein.
XX
OS Homo sapiens.
XX
FN WO200190304-A2.
XX
PD 29-NOV-2001.
XX
PF 18-MAY-2001; 2001WO-US016450.
XX
PR 19-MAY-2000; 2000US-0205515P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Birse CE, Rosen CA;
XX
DR WPI; 2002-122018/16.
XX
DR N-PSDB; ABL90662.
XX
PT Novel 1405 isolated polypeptides, useful for diagnosis, treatment and
PT prevention of neural, immune system, muscular, reproductive,
PT gastrointestinal, pulmonary, cardiovascular, renal and proliferative
PT disorders.
XX
PS Claim 11; SEQ ID NO 2629; 2081pp + Sequence Listing; English.
XX
CC The invention relates to novel genes (ABU89449-ABU90853) and proteins
CC (ABU89040-ABU90444) useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy. The genes are
CC isolated from a range of human tissues disclosed in the specification.
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful in
CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and
CC ovarian cancer and other cancers of the adrenal gland, bone, bone marrow,
CC breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune
CC disorders e.g. Addison's disease, allergies, autoimmune haemolytic
CC anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,

CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)
CC cardiovascular disorders such as myocardial ischaemias; (d) wound healing
CC ; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)
CC infectious diseases such as viral, bacterial, fungal and parasitic
CC infections. Note: the sequence data for this patent did not form part of
CC the printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 279 AA;

Query Match 14.1%; Score 1215; DB 5; Length 279;
Best Local Similarity 88.4%; Pred. No. 6.5e-76;
Matches 245; Conservative 0; Mismatches 0; Indels 32; Gaps 1;

Qy 552 MECDDLKELPEPTPVKTRLPPEIFGDMVLEPLNFAFGELFDLQDEPDGVTLEVLREAL 611
Db 1 MECDDLKELPEPTPVKTRLPPEIFGDMVLEPLNFAFGELFDLQDEPDGVTLEVLREAL 60

Qy 612 VGNDSEGPLCELLFFFLTAIFQAIAEEEEVEAKEQLTDADTK----- 653
Db 61 VGNDSEGPLCELLFFFLTAIFQAIAEEEEVEAKEQLTDADTK----- 120

Qy 654 -----GCSLSKSLDSDCTLSEILRLHILASGADVTSANAKRYQKRGGFDA 699
Db 121 AVASLAAAWPQLHQGCSLSKSLDSDCTLSEILRLHILASGADVTSANAKRYQKRGGFDA 180

Qy 700 TDDACMELRLSNPSLVKLSSTSVYDLTPGKMKILHALCGKLLTLVSTRDFIEDYVDIL 759
Db 181 TDDACMELRLSNPSLVKLSSTSVYDLTPGKMKILHALCGKLLTLVSTRDFIEDYVDIL 240

Qy 760 RQAKQEFRELKAEQHRKEREAEAAARIRKREKLEKEQ 796
Db 241 RQAKQEFRELKAEQHRKEREAEAAARIRKREKLEKEQ 277

RESULT 8
ADQ20910
ID ADQ20910 standard; protein; 1483 AA.
XX
AC ADQ20910;
XX
DT 26-AUG-2004 (first entry)
XX
DE Human soft tissue sarcoma-upregulated protein - SEQ ID 3730.
XX
KW soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human.
XX
OS Homo sapiens.
XX
FN WO2004048938-A2.
XX
PD 10-JUN-2004.
XX
PF 26-NOV-2003; 2003WO-US038193.
XX
PR 26-NOV-2002; 2002US-0429739P.
XX
PA (PROT-) PROTEIN DESIGN LABS INC.
XX
PI Aziz N, Ginsburg WM, Zlotnik A;
XX
PI WPI; 2004-441208/41.
XX
DR
XX
PT Early detection of soft tissue sarcoma comprises determining expression
PT of a gene in a first soft tissue sample and a normal soft tissue sample
PT and comparing the gene expression, also useful in treating soft tissue
PT sarcoma.
XX
PS Example 2; SEQ ID NO 3730; 210pp; English.
XX
CC The invention relates to a novel method for detecting soft tissue sarcoma
CC which comprises obtaining a first soft tissue sample from an individual
CC and a normal soft tissue sample from the same or different individual,

CC determining the expression of a gene in both samples and comparing the
 CC expression of the gene in both soft tissue samples, where a higher level
 CC of protein expression in the first soft tissue sample indicates the
 CC presence of soft tissue sarcoma. The method of the invention has
 CC cytostatic applications and may be useful for detecting soft tissue
 CC sarcoma, possibly via gene therapy or vaccine production. The nucleic
 CC acid sequences may be useful in diagnostic and screening applications.
 CC The current sequence is that of a human soft tissue sarcoma-upregulated
 CC protein of the invention. The current sequence is not shown within the
 CC specification per se but was submitted in CD format by the inventor.
 XX
 XX Sequence 1483 AA;

Query Match 11.0%; Score 945.5; DB 8; Length 1483;
 Best Local Similarity 23.3%; Pred. No. 4.8e-56;
 .Matches 400; Conservative 262; Mismatches 522; Indels 533; Gaps 67;

QY 152 PLLRRKPFVRQPPADLRDEBEVEYCKVTNETFRHYDDPFERTILCNLSVNCATVGRPG 211
 DB 3 PLGRKPPPLVRP---LPGEELFTPIPTQEAFTREYEARELERYSERIWTCKSTGSSQ 59
 QY 212 LTYQALESEKKARQL-OSFF---EPLIIPVLYLTSLTHRSRLHEICDDIPAYYKDRYF 267
 DB 60 LTHKAWEEEQEVAELLKEEFPAWYKVLVENVH---HNTASLEKLVDTAWLEIWTKYA 115
 QY 268 VBETVEVIRNNGARLOCTILEVLP-----PS-----HONGFANGHYN 304
 DB 116 VGEEDCFEYGEKMLKVIVIEHPLEKVDDEATEKKSOGACDPSDSENSQIAQDHQK 175
 QY 305 SVDGITIIISDSDSETQS-----CSFQNGKKDAIDPLL-FKYKQPTKKELH 352
 DB 176 K---ETVVKEDGRRSINDRARRSPKLPSTLKKGERKWPAPKPLPHKYDV-----KLQN 228
 QY 353 ESAI---VKATOISRRKLFSDKLFKLCQRC-----EPOEGVIK---IKASSLTVK 400
 DB 229 EDKIISNPADSLIRTERPPNKEIVRYFIRHNALRAGTCENAPWVVEDELVKYSLPS-- 286
 QY 401 IAEQDSVFPPDDPTFTFSPANRRG-----RPPKRIHSQEDN----- 440
 DB 287 ----KPSDFLLDPYKMTLNPSTKTKNGTSPDRKPSKK---SKTONSSLSPLNPKLWCH 339
 QY 441 VANKOTLA-----SYRSKATKER-----D 459
 DB 340 VHLKXSLSGSLPKVKNKSNKSPESBLEEMMQMSPNKLHTNFHIPKKGPPAKPGKHS 399
 QY 460 KLLK-----QEWKSLAFKAKLKR----- 479
 DB 400 KPLKAKGRSKGILNGQKSTGNSKSPKGLKTKTKQKQNTLDMAKGTQKMTAPRNSGG 459
 QY 480 -----EKADALEAKKKEKEDKEKR-----EELK 503
 DB 460 TPRTSSKPHKLPPLPAALHLIAYYKENDKREDKRSALSCVISKTARLLSSDRAPLPEELR 519
 QY 504 KIVEE-----RLKKEEGERLKVREKEREKURE--EKRKVYVYLKQWSPR- 549
 DB 520 SLVQKRYELLEHKRWASMESEQRKYLEKKEELKXKLKEKAKERREKEMLERLEKQKR 579
 QY 550 -EDMECDLDELPEPTPVKT--RLPEIFGDALMVFLEFLNAGELFDLQDEPDGVTLEV 606
 DB 580 YEDQELTG-KNLPAPRLVDTPGLNLTFLGDMVWVFEFLSCYSGLLLPDAQIP--ITAVS 636
 QY 607 LBEALVGNDSGELPLCELLFFFTAI FQAIABE--BEEVAKEQLTDADTKGCSLSKLSDLS 664
 DB 637 LMEAL-SADKGG-----FLYLNRLVILLQTLQDEIAE---DYGELGKMLSEIPLTL 685
 QY 665 CTLISILRLHILASADVTISANAKRYQKRGFDATDDACMELRLSNFSLVKLSSTSVY 724
 DB 686 HVSLSVLRLCLRRSQVSESGSDTD-----DNKDSAAFEDNEVDQEFLEKLETSFF 738
 QY 725 DLTPECKMKLILHALCGKLLTLVSTRDFIEDYVDILRQAKQEPRELKAEQHRKEREAAAR 784
 DB 739 ELTSEKQLILTALCHRLIMTVSVQD-----HMETROQMGAEE 775

QY 785 IRKRKEEKLKEQEQKMKQKQKLEKDE---QRNSTADISIGEEREDFDTSIESKDTQOK 841
 DB 776 LMKRELAVLKEENDKKRAEKQKQKREMEAKNKENGKVENGLKXTDRK-----KETV 825
 QY 842 ELDDQDMETEDDDPGSHKGRG---RGKRGQNGKFEFTROEQINCVTRRELLTADBEELK 897
 DB 836 KFEPOVDTAEADMISAVKSRRLAIQAK-----KEREIQE-----REMKVKLBRQA-E 872
 QY 898 QEHQKKEKELLEK-IQSAIACTNIP---PLGRDRMYRRYWF-PSIRGLFIEEDYSGIT 951
 DB 873 EERIRKHAARAKAFQEGIAKAKLVMRRTPIGTDRHNHRYMLFSDVFGCLFIEK---GMV 929
 QY 952 EDMLLPRPSSFNQVSDPOVSTYKTEGP-----LMSESTSNIDQG-PRDHSVOL 1000
 DB 930 HDSI-----DYRFNHCHDKHTVS---GDEYCPRSKANKLGNKSNMTQHGTAATEVAVET 981
 QY 1001 PKPVHKPNRMCFYSCOEOLDOLIEALNSRGHRESALKETLLOEKSRIQAQLARFSE--BK 1058
 DB 982 TTPQGGQMLFCDQSKDELBNCLHPQGIRESOLKERL--EK-----RYQDIHS 1031
 QY 1059 FHFSDKPOQPSKPTYSRGRSSNAYDPSQWCAEQLELRDLDFLLDIEDRIYQGTILGAIKV 1118
 DB 1032 IHLARKPNLGLK-----SCDGNQ-----ELLNFLRSDLIEVATRLQKGLGYVEE 1076
 QY 1119 TDRHWRSALESRGYELLSEENKENGIIKTVDNEVEEMEIDEQTKVIVK--DRLLGIKTE 1176
 DB 1077 T-----SEPEA---RVISLEKL-----KDFGECVIALQASVKKFKLQGFMAKQK 1118
 QY 1177 TPSTVSTNASTPQSIVSSVVHYLAMALFOEQIEERRFLKAPLDASDSGRSYKTVLDRRE 1236
 DB 1119 RRLQSEDSATEEVE-----EKQWVEEAKVAS-----ALEKWT 1154
 QY 1237 SLLSASLSQVFLHLSTLDRSVIWSILNARCKIKRKGDAENMVLCDGDRGHHTYCV 1296
 DB 1155 AIREAQTFSRMHLVLLGMLDACIKWMSAENARCKVCKKGGDDKLILCDECNKAHLFCL 1214
 QY 1297 RPKLKTYPGQWPCPCRPKORCRLSPRQPSLESDVEDSMGGEDEVDGDEEGOS 1356
 DB 1215 RPALYEVDPGEWQCPACQAPATARRNSGRNYTEBSASEDSEDESDEEE---EEEBEE 1270
 QY 1357 EEEVEVEQDEDDQSEEEVSLPKRGRPVRL-PVKT--RGKLSFSFSSRQGOQEGRYPS 1414
 DB 1271 EEEYEV-----AGLRLPRKTIRGHSHVIPPAAARSGRRPGKXPH 1310
 QY 1415 RSQOSTPVTSSKTGRSLRKINSAPPTETKSLRIASRSTRHSHGLQADVVELLSPPR 1474
 DB 1311 STRSQPK-----APPVD-----DAEVDL----- 1330
 QY 1475 KRRGRKSANTPENSPNPNFRVIATKSSQSRSVNIASKLSLOESKRRCKRQSPPEP 1534
 DB 1331 -----VLQTKSSRRQSL----- 1343
 QY 1535 SPVTLGRSSRGQGVHLSAFEQVLVVLVPHDDSWPFLKLVSKIQVDPYDVIKKPIAL 1594
 DB 1344 -----ELQKCEILHKIVYRFSWPFEPVTRDEAEDYDVIHPMDP 1386
 QY 1595 NIIEKVNKCYKLASEFIDDIELMFSCFYNPNPT 1631
 DB 1387 QTVQNKSCGSYRSVQEFELTDMKQVFTNAEYVNCGRS 1423
 RESULT 9
 ID ADF69140
 XX ADF69140 standard; protein; 1483 AA.
 AC ADF69140;
 XX
 DT 12-FEB-2004 (first entry)
 XX Human MP53 protein sequence SEQ ID NO:110.
 DE p53 pathway modulating agent; MP53; p53 modulator; cytostatic;
 KW gene therapy; cancer; human.

XX Homo sapiens.
 OS WO2003083047-A2.
 PN 09-OCT-2003.
 PD 28-FEB-2003; 2003WO-US006025.
 PF 01-MAR-2002; 2002US-0361196P.
 PR (EXEL-) EXELIXIS INC.
 PA Belvin M, Francis-Lang H, Friedman L, Plowman GD, Heuer TS, Li D;
 PI Funke R;
 PT
 PI
 XX
 DR WPI; 2003-812540/76.
 DR N-PSDB; ADF69196.
 XX
 PT Identifying a candidate p53 pathway modulating agent for treating e.g.,
 PT cancer by contacting an assay system comprising a MP53 polypeptide or
 PT nucleic acid with a test agent and detecting a test agent-biased
 PT activity.
 XX
 PS Example; SEQ ID NO 110; 406pp; English.
 XX
 CC The present invention describes a method for identifying a candidate p53
 CC pathway modulating agent, which comprises: (a) providing an assay system
 CC comprising a MP53 (modulator of p53) polypeptide or nucleic acid or its
 CC fragment or derivative; (b) contacting the assay system with a test agent
 CC under conditions where the system provides a reference activity except in
 CC the presence of the test agent; and (c) detecting a test agent-biased
 CC activity, where a difference between the test agent-biased activity and
 CC the reference activity identifies the test agent as a candidate p53
 CC pathway modulating agent. Also described: (1) modulating the p53 pathway
 CC of a cell; (2) modulating the p53 pathway in a mammalian cell; and (3)
 CC diagnosing a disease in a patient. MP53 has cytostatic activity, and can
 CC be used in gene therapy. The method is useful for identifying a candidate
 CC p53 pathway modulating agent for preparing a composition for diagnosing
 CC or treating e.g., cancer. The present sequence represents a human MP53
 CC protein, which is used in the exemplification of the present invention.
 XX
 SQ Sequence 1483 AA;
 Query Match 10.9%; Score 938.5; DB 7; Length 1483;
 Best Local Similarity 23.2%; Pred. No. 1.5e-55;
 Matches 399; Conservative 262; Mismatches 523; Indels 533; Gaps 67;
 QY 152 PLLHRKPFVRQKPPADLPDEBFYCKVTNFIHRHYDDFFERTILCNLSLVWSCAVTGRPG 211
 DB 3 PLLGRKPPLVKP---LPGEPPLFTIPIHTQEAFTREEYEARELRYSERIWTCKSTGSSQ 59
 QY 212 LTYQALSEKKARQL--QSPF---EPLIIPVLYLTSLTHRSRLHEICDDIFAVVKORYF 267
 DB 60 LTHKEAWEEQVABLLKEEPFAPWEKLVLEMVH---HNTASLEKLVDTAWLEIMTKYA 115
 QY 268 VEETVEVIRNNGARLOCTILEVLP-----PS-----HONGFANGHVN 304
 DB 116 VGEECDFFVGEKMKLVKVIKIHLEKVDTEATEKKSDGACDSPSKENSSQIAQDHQK 175
 QY 305 SVDGETIISDSDSETQS-----CSFQNGKKKDAIDPLL-FKYKVQPTKKELH 352
 DB 176 K---ETVVKEDEGRRESINDRARRSPKLPITSLKGERKWAFFPLPHKYDV-----KLON 228
 QY 353 ESAI---VKATOISRKKHFRDCLKFLKQHC-----EPQGVK---TKASSLSYTK 400
 DB 229 EDKIISNPADSLIRTERPPNKEIVRYFIRHNALRAGTGENAPWVVEDELVKYSLPS-- 286
 QY 401 IAEQDSFYFPDPTTFIFSPANRRG-----RPPKIHISQEDN----- 440
 DB 287 ----KFSDFLLDPYKMTLNPSKTKNTGSPDRKPSKK---SKTDNSSLSPNPKLWCH 339
 QY 441 VANKQTLA-----SYRSKATKER-----D 459

DB 340 VHLKSLSGSPLKVKNSKNSKSPPEHLEEMKOWSPNKLHTNFHIPPKKGPAPKPKGKHS 399
 QY 460 KLIK-----QEEKSLAFKAKLKR----- 479
 DB 400 KPLKAGRSKILNGQKSTGNSKSPKGLKTPKTMKQMTLLDMAKTQKWTRAPRNSGG 459
 QY 480 -----EKADALEAKKEDKEKR-----BELK 503
 DB 460 TPRTSSXPHKHLPPAALHLIAYKENKDRDEKKSALSCVISKTARLILSSSEDRARLPBLR 519
 QY 504 KIVEE-----RLKKKEKERLKVREKEREKJRE--EKRYKVEYLKOWSKPR- 549
 DB 520 SLVKRYELLEHKRWASMEEQKEYLKKREBELKKLKEKAKERKEKEMLERLEKQKR 579
 QY 550 -EDMECDLDELPEPTPVKT--RLPPEIFGDALMVLEFLNAGFELFDQEPFGVTVLEV 606
 DB 580 YEDQELTG--KNLPAPRLVDTPEGLPNTLFGDVAMVVEFLSCYSGLLLPDAQYP--ITAVS 636
 QY 607 LEBALVGNDSGELCELLFFFLTAIFQAIABE--EEBVAKEQLTDADTKGCSLKSLLDS 664
 DB 637 LMEAL--SADKGG-----FLYLNRLVILLOTLQDEIAB-----DYELGNKLSIPLTL 685
 QY 665 CTLSEILRLHLASGADVTSANAKRYQKRGGFATDDACMELRLSNPSLVKLSSTSVY 724
 DB 686 HVSVELVRLCLRRSDVQEESESGDTD-----DNKDSAAFEDNEVDQEFLEKLETSEFF 738
 QY 725 DLTPEGMKILHALCGKLLTVSTRDFIEDYVDILROAKQEFRELKABQHKREEREAAR 784
 DB 739 ELTSEELQLITLALCHILMTYSVQD-----HMETRQOQMSAE 775
 QY 785 IRKKEEKLKEQEKMEKQEKLEKEDB---QRNSTADISIGEEREDFDTIESKDTQOK 841
 DB 776 LMKELAVLKEENDKRAEKQKREKAKKENGKVENGLKTDK-----KEIV 825
 QY 842 ELDDQMTFEDDDPGSHKRG---RGKRGQNGKPEFTROEQINCVTRELLTADDEEALK 897
 DB 826 KFEPQVDTAEADMTISAVKSRLLAIQAK-----KEREIQE-----REMKVKLERQA-E 872
 QY 898 QEHORKEKLELEK-IQSAIACTNIF-----PLGRDMYRRYWIIP-PSIPGLFIEDYDGLT 951
 DB 873 EERIRKHAEEKAFQEGIAKAKLVMMRTPIGTDRNHNRYLWFSDEVPGLFIEK---GWV 929
 QY 952 EDMLLPRPSSFONNVQSDPQVSTKTGEP-----LMSESTSNIDQ--PRDHSVQ 1000
 DB 930 HDSI-----DYRFNWHCKDHTVS---GDEYCYPRSKKANLGKASMTQHGTADEVAVET 981
 QY 1001 PKPVHKNRWCYSSCEQLDOLIEALNSRGHRSALKETLLQEKSRICAQLARFSE--EK 1058
 DB 982 TTPKQGNLWFLCDSQKELDELLNCLHPQIGRESOLKRL--EK-----RYQDIHS 1031
 QY 1059 FHFSKDPQPSKPTYSGRSSNAYDPSOMCAEQLELRDLFLDIEDRIYOGTLGAIKV 1118
 DB 1032 IHLARKENLGLK-----SCDGNQ-----ELLNFLRSDLTIEVATLQKGLGVVEE 1076
 QY 1119 TDRHIWESALSGRYELLSBENKENGIIKTWNEDVEEIMEIDQPKVIK--DRLLGIKTE 1176
 DB 1077 T-----SEFEA---RVISLEKL-----KDFGECVIALQASVKKFKLOGFMAKQK 1118
 QY 1177 TPSTVSTNASTPOSVSVVHYLAMALFOIBQIERRELFKAPLDASDSGRSYKTVLDWRRE 1236
 DB 1119 RRKLQESDAKTEBDE-----EKKWVEEAKVAS-----ALEKWK 1154
 QY 1237 SLSSASLSQVFLHSLTDSRVWSKSIINARCKICRKGDAENMWCDGCDRGHHTYCV 1296
 DB 1155 AIREAQTFSRMHVLLGMDACIKWDMGAENARCKVCRKGGEDDKLILCDECNKAFHLFCL 1214
 QY 1297 RPKLKTVPEGDFWCFECPKQRCRLFRQPSLESDVEDSDMGDDDEVDGDEEESGOS 1356
 DB 1215 RPALYEPDGEWQCPAQCPATARNSGRNVTESESASEDDEDEDEE-----EEEE 1270
 QY 1357 EEEYEYEDDDSDQEEEEVSLPKRGRPVRL--PVKT--RGKLSSSFSFGQQQPGRYPS 1414

Db 1271 EEDYEV-----AGLRPRKTRGKSHVIPPAAARGRRGKKPH 1310
Qy 1415 RSQOSTPKTVSSKTRSLRKINSAPPTETKSLRIASRSTRSHGFLQADVVELLSPRR 1474
Db 1311 STRRSQPK-----APPVD-----DAEVEL----- 1330
Qy 1475 KRGKRSANTPSPNPNFVIAIKSEQSRSVNIASKLSQSESKRRCKRQSP 1534
Db 1331 -----VLQTKSSRRQSL----- 1343
Qy 1535 SPVTLGRSSRGQGVHLSAFLQVLVHDDSWPFLKLVSKIQVDPYDIIKKPIAL 1594
Db 1344 -----ELQCEBILHIVKIRFSWPFRPDRDEADYDVIITHPMDF 1386
Qy 1595 NIIREKVNKCEYKLASEFIDDIELMFNSCFEYNPRNT 1631
Db 1387 QTVQNKCSGYSRSQVEFLTMKQVFTNAEVNCRGS 1423

RESULT 10
ID AAW81173
XX AAW81173 standard; protein; 1531 AA.
AC AAW81173;
XX
DT 05-MAY-1999 (first entry)
XX
DE Human BAZ1-beta protein #2.
XX
KW Transcriptional regulator; transcription; BAZ1-alpha; bromodomain; BAZ;
KW atypical zinc finger; testis; human; tumour; BAZ1-beta; BAZ2-alpha; drug;
KW BAZ2-beta; treatment; cancer; proliferative disorder; screening.
XX
OS Homo sapiens.
XX
PN WO9847920-A1.
XX
PD 29-OCT-1998.
XX
PF 17-APR-1998; 98WO-JP001783.
XX
PR 18-APR-1997; 97JP-00116570.
PR 24-OCT-1997; 97JP-00310027.
XX
PA (CHUS) CHUGAI RES INST MOLECULAR MEDICINE INC.
XX
PI Jones MH;
XX
DR WPI; 1998-583603/49.
DR N-PSDB; AAV68409.
XX

PT Transcriptional regulator gene family containing bromodomain - may be
PT expressed in testis tissue and is useful for treatment of cancer and
PT other proliferative disorders.
XX
PS Claim 1; Page 145-157; 187pp; Japanese.
XX
CC This sequence represents the human BAZ1-beta protein, a member of a
CC family of transcriptional regulator genes containing a bromodomain (BAZ,
CC Bromodomain with Atypical Zinc finger) which are expressed specifically
CC in testis tissue and also in certain tumour lines. Transgenic cells may
CC be used for the preparation of the BAZ1-alpha, BAZ1-beta, BAZ2-alpha and
CC BAZ2-beta proteins. These proteins can be used in the treatment of cancer
CC and other proliferative disorders, and in screening of compounds for
CC their binding ability to the expression products (e.g. for use as drugs
CC by modulation of transcriptional regulation)
XX
SQ Sequence 1531 AA;

Query Match 10.8%; Score 930.5; DB 2; Length 1531;
Best Local Similarity 23.2%; Pred. No. 5.6e-55;
Matches 399; Conservative 263; Mismatches 526; Indels 529; Gaps 68;

Qy 152 PLLHRKPVROKPADLRDPDEVFY-CKVWNEI FRHYDDFFERTILCNLSVWSCAVTGRP 210
Db 3 PLLGRKP-----PLVNPPLGSEPFPTPHQTQAFRTREYEARELERYSERIWTCKSGSS 58
Qy 211 GLTQEALESKKARQNL-QSFP---EPLIIPVLYLSLTHRSRLHEICDDIFAVVKDORY 266
Db 59 QLTHKEAWEEQOEVAELLKEEFPWYKLVLEMVH---HNTASLEKLVDTAWLEIMTKY 114
Qy 267 FVEETVEVIRNNGARLOCTILEVLP-----PS-----HONGFANGHV 303
Db 115 AVGEEDCEVEGCKMVKVIVKIHPLEKVDDEATEKSDGACDSPSSDKENSSQIAOHDQ 174
Qy 304 NSVDGETIIISDDSETQS-----CSFQNGKKDAIDPLL-FKYKVQPTKKEL 351
Db 175 KK---ETVWKEDEGRRESINDRARRSPRKLTSLKGERKWAQPKFLPHKYDV-----KLQ 227
Qy 352 HESAI---VKATOISRRKHLFSRDKLKFLKQHC-----EPQGVK---IKASSLSY 399
Db 228 NEDKIISNVPADSLIRTERPPNKEIVRYFIRHNALRAGTGENAPWVVEDELVKYKSLPS- 286
Qy 400 KIAEQDFSYPFPDPPPTFIFSPANRRG-----RPPKRIHISOEDN----- 440
Db 287 -----KFSDFLLDPYKYMTLNPSTYKNTGSPDRKPSK---SKTDNSSLSPLNPKLWC 338
Qy 441 -VANKQTLA-----SYRSKATKER----- 458
Db 339 HVHLKXSLSGSPLKVXNSKNSKSPPEHL EEMMKWSPNKLHTNFHPIPKGPPAKKPGKHS 398
Qy 459 DXLJK-----OEEMKSLAFKAKLKR----- 479
Db 399 DKPLKAKGRSGKILGQKSTGNSKSPKGLKTPKTKMQLTLDMAKGTQXWTRAPNSG 458
Qy 480 -----EKADALEAKKKEKDEKKK-----EEL 502
Db 459 GTPRTSSKPHKHLPPAALHLIAYYKENKDRKRSALSCVTSKTAARLLSSDDRARLPDEL 518
Qy 503 KKIYEEE-----RLKKKEERLKVVEREKREKLRE---EKRYVEYLYLKQWSKPR 549
Db 519 RSLVQKRYELLEHKKRWASMSSEQRKEYLVKKRBEELKKLKEKAKERREKEMLERLEKQK 578
Qy 550 --EDMECDLKLPEPTPVKT--RLPPEIFGDALMVLEFLNAGFELFDLQDEFFDGVTL 605
Db 579 RYEDQELTG-KNLPAPFLVDTPEGLPNTLFGDVAMVVEFLSCYSGLLLPDAQYP--ITAV 635
Qy 606 VLBEALVGNDSG--PLCELLFPFLTAIFAIAEEREEVAKEQLTADTKCSLSKSLDLD 663
Db 636 SLMEAL-SADKGGFLYLNRLVILLQTLQLL--QDEIAB---DYGELGNKLEIPLT 688
Qy 664 SCTLSEILRLHLASGADVTSANAKRYQKRGSGFDATDDACMELRLSNPSLVKLSSTSV 723
Db 689 LHSVSELVRLCLRSVDQSESGSDTD-----DNKDSAAFEDNEVDQEFLEKLETSEF 741
Qy 724 YDLTPGKMKILHALCGKLLTLVSTROFIEDYVDILRQAKQSFRELKAEQHRKEREAAA 783
Db 742 FELTSEEKQLILTALCHRIILMTYSVD-----HMETROQMSA 778
Qy 784 RIRKKEEKLQEQKMKQKKEKLEDEQBNSTADISIGEEEREDFTSIESKTEQK-- 841
Db 779 ELWKERLAVLKEENDKKRAEKQKKEAKN-----KENGKVENGLGKTDRKKRIV 829
Qy 842 ELDDQMTFEDDDPGSHKRG---RGKRGONGKFEFTROBQINCVCVTRRELLTADDEBALK 897
Db 830 KFEPOVDTAEADMISAVKSRLLAIOAK-----KEREIQE-----REMKVKLERQA-E 876
Qy 898 QEHORKEKLELEK-IQSAIACNTIF---PLGRDMYRRYWI-F-PSIPGFIPIEDYSGLT 951
Db 877 EERIRKHAATAEAFQEGIAKAKLVMRRTPTGTRNHNRYWLFSDVEVGLFIEK---GWV 933
Qy 952 EDMLLPRSPFQNNVQSDPQVSTKTGEP-----LMSESTSNIDQG-PRDHSVOL 1000
Db 934 HDST-----DYRFNHHCXDHVS---GDEDYCPRSKXANLGNKASMTQHGTA TEVAVET 985
Qy 1001 PKPVHKPNRWCFYSSCQLDQLIBALNSRGHRESALKETTLQEKSRICAQLARFSE--EK 1058

```
Db 986 TTPKQGNLWFLCDSQKELDELLNCLHPQIGRESQLKERL--EK-----RYQDIHS 1035
Qy 1059 PHFSDKPPQDSKPTYSRGRSSNAYDPQMCBAEQLELRFLDIEDRIYQGTILGAIV 1118
Db 1036 IHLAKPNLGLK-----SCDGNQ-----ELLNFRDLIEVATRLQKGLGYVEE 1080
Qy 1119 TDRHWRSALESGRYELLSEENKENGIIKTVDNEVEEIDBQTKVIVK--DRLLGIKTE 1176
Db 1081 T-----SEFPA---RVLSLEKL-----KFGECVIALQASVKKFLQGFAPKQK 1122
Qy 1177 TPSTVSTNASTPQSVSSVHYLAMALFOIEQIERFLKAPLDASDSGRSKYTKVLDRWRE 1236
Db 1123 RKQLQSEDSAKTEEVD-----EKKMVEAKVAS-----ALEKWK 1158
Qy 1237 SLLSASLSQVFLHLSTLDRSVIWSKSLNARCKIKRKGDAENWVLCDGCDRGHHTYCV 1296
Db 1159 AIREAQTFSRMHVLLGMLDACIKWDMSAENARCKVCPKKGGDDKLILCDECNKAFHLFCL 1218
Qy 1297 RPKLKTVPEDGWFCEPCRPKQRCRLSPQRPSLESDEDEDSMGDEVDGDEEGOS 1356
Db 1219 RPYLYEVPDGEWQCPACQATARRNSRGRNTEESASEDSDEDEDEE-----EBEEEE 1274
Qy 1357 EEEYEVEQDEDDSOEEBEVSILPKRGRPVRL-FVKT-RGKLSSFSRSGQQQBPGRYPS 1414
Db 1275 EEEDEYEV-----AGLRLRPKRTIRGKHSVIPPAAARSRRPGKKPH 1314
Qy 1415 RSQQTPTKTVSSKTGRSLRKINGSAPPTETKSLRIASRSTRSHSGPLQADVVELLSPRR 1474
Db 1315 STRRSQPK-----APPVD-----DAEVDL----- 1334
Qy 1475 KRRGRKSANNTPENSPNPNFRVIATKSSQSRSVNIASKLSLQSESKRCRKRQSEPEP 1534
Db 1335 -----VLQTKRSRRQSL----- 1347
Qy 1535 SPVTILGRRSGRQGVHLSAFAEQVLVVELVRHDDSWPFLKLVSKIQVPDYDIIKKPIAL 1594
Db 1348 -----ELQCEELIKHVIKYRFSWPPEPVRDEAEDYDVITHPMDF 1390
Qy 1595 NIIREKVNKCEYKLASEFIDDIELMFSCNCFYNPNT 1631
Db 1391 QTVQNKCGSVRSYQVEFLTMKQVFTNAEVNCRGS 1427

RESULT 11
ID AAW81172 standard; protein; 1527 AA.
XX AAW81172;
AC AAW81172;
DT 05-MAY-1999 (first entry)
DE Human BAZ1-beta protein #1.
XX
KW Transcriptional regulator; transcription; BAZ1-alpha; bromodomain; BAZ1;
KW atypical zinc finger; testis; human; tumour; BAZ1-beta; BAZ2-alpha; drug;
KW BAZ2-beta; treatment; cancer; proliferative disorder; screening.
XX
OS Homo sapiens.
XX WO9847920-A1.
XX
PD 29-OCT-1998.
XX
PF 17-APR-1998; 98WO-JP001783.
XX
PR 18-APR-1997; 97JP-00116570.
XX
PR 24-OCT-1997; 97JP-00310027.
XX
PA (CHUS ) CHUGAI RES INST MOLECULAR MEDICINE INC.
XX Jones MH;
PI
XX
```

```
DR WPI; 1998-583603/49.
XX N-PSDB; AAV68408.
PT Transcriptional regulator gene family containing bromodomain - may be
PT expressed in testis tissue and is useful for treatment of cancer and
PT other proliferative disorders.
XX Claim 1; Page 125-137; 187pp; Japanese.
XX
CC This sequence represents the human BAZ1-beta protein, a member of a
CC family of transcriptional regulator genes containing a bromodomain (BAZ,
CC Bromodomain with Atypical Zinc finger) which are expressed specifically
CC in testis tissue and also in certain tumour lines. Transgenic cells may
CC be used for the preparation of the BAZ1-alpha, BAZ1-beta, BAZ2-alpha and
CC BAZ2-beta proteins. These proteins can be used in the treatment of cancer
CC and other proliferative disorders, and in screening of compounds for
CC their binding ability to the expression products (e.g. for use as drugs
CC by modulation of transcriptional regulation)
XX
SQ Sequence 1527 AA;
Query Match 10.8%; Score 929.5; DB 2; Length 1527;
Best Local Similarity 23.1%; Pred. No. 6.5e-55;
Matches 397; Conservative 264; Mismatches 523; Indels 533; Gaps 68;
Qy 152 PLLHRKPFVRQKPPADLRPDPEVFY-CKVTNEIFRHYDDFFERTILCNLSVWSCAVTGRP 210
Db 3 PLGRKPF-----PLVNPDPGEEFFPTIPHQEAFRTREYEARELRYSERIWTCKSTGSS 58
Qy 211 GLTYQEALESSEKARQNL-QSFP---EPLIIPVLYLTSLTHSRHLHEICDDIFAYVKDXY 266
Db 59 QLTHKEAWEEBQEAELKEEFPAWYKLVLEMVH---HNTASLEKLVDTAWLEIMTKY 114
Qy 267 FVEETVEVIRNNGARLOCTILEVLP-----PS---HQNGFANGHV 303
Db 115 AVGECDFEYKGMKLVKIVKHPLEKVDDEATEKSDGACDPSPSDKENSSQIAQDHQ 174
Qy 304 NSVDGETIISDSDDSTQS-----CSFONGKKKDAIDPLL-FYKVKQPTKEL 351
Db 175 KK---ETVVEDEGRRESINDRARRSPRKLTSLKGERKWPAPKFLPHKYDV---KLQ 227
Qy 352 HESAI---VKATQISRKHFLSRDKLKLFLKQHC-----EPOEGVIK---IKASSISTY 399
Db 228 NEDKIISNVFADSLIRTPPENKEIVRYFIRHNALRAGTGENAPWVVEDELVKYSLPS- 286
Qy 400 KIAEQDFSYFPDDPPTFFISPANRRG-----RPPKRIHISOEDN----- 440
Db 287 ----KFSDFLLDPKYMTILNPSTKRKNTGSPDRKPSK---SKTDNSSLSLSPNPKLWC 338
Qy 441 -VANKOTLA-----SYRSKATKER----- 458
Db 339 HVHLKSLSGSPLKVNKSNKSPSEEHLEEMKMSPNKLTNPHIPKGPAPKPKGHS 398
Qy 459 DKLIK-----QEEKSLAFKAKLKR----- 479
Db 399 DKPLAKGRSGKILNGKSTGNSKSPKGLKTPTKMKQMTLLDMAKGTQKTRAPRNSG 458
Qy 480 -----EKADALEAKKKEKDEKKEK-----EEL 502
Db 459 GTPRTSSKPHKHLPPAALHLIAVYKKNKDKREDKRSALSVCVISKVARTLLSSEDRARLPEEL 518
Qy 503 KKIVEEB-----RLKKKEEKERLVEREKERKLE--EKRYVEYLKQWSKPR 549
Db 519 RSLVKRYELLEHKKRWASMSSEORKEYLKKRELKKLKKERKEKREKEMLERLEKQK 578
Qy 550 --EDMECDLKEPEPTFVKT--RLPPEIFGDALMWLEFLNAGFELFDQDEPDGVTL 605
Db 579 RYEDQELTG-KNLPAFLVDTPEGLPNTLFGDVAMVVEFLSCYSGLLLPDAQYP--ITAV 635
Qy 606 VLEBALVNDSEGLPCELLFFFLTAIQAIAEB--EEEVAKEQLTDADTKGCSLKSJLDD 663
Db 636 SLMEAL-SADKGG-----FLYLNRLVLVLIQTLLQDEIAE-----DYGELGMKLSIPLT 684
```



```
QY 664 SCTLSEILRLHLASGADVTSANAKYVQKRGFFDATDDACMELSLNPSLVKLLSSTSV 723
Db LHSVSELVRLCURSDVQEESEGSTD-----DNKSAAPFEDNEVQDFLEKLETFSEF 737
QY 724 YDLTPGCKMKILHALCGKLLTLVSTRDFTIEDYVDILRQAKQFRELKAEQHKREERAAA 783
Db FELTSEEKLIULTALCHRLMTYSVD-----HMETRQOMSA 774
QY 784 RIRKKEKLPQEQKMEKQKLBEDBORNSTADISIEBEREDFDTSIESKQTEQK-- 841
Db ELKELRLAVLEBKAEKQKREMEAKN-----KENGKVENGLGKTRDKRIV 825
QY 842 ELDDQMFTEDEDDPSHKRGR-----RGKRGQGFKEFTROEQINCVCVRELLTADREALK 897
Db KPEQVDTEAEDMISAVKSRRLLAIOAK-----XEREIQE-----REMKVLERQA-E 872
QY 898 OBHOREKELEK-TQSAIACTNIF-----PLGRDMYRYWIF-PSIPGLFTEEDYSGLT 951
Db EERIRKHAAKAEKQEGIAKAKLVMRPTIGTDNRHRYWLFSEVFGLFIEK-----GWV 929
QY 952 EMLLPRPSSFNQNVOSQDPQVSTKTGEP-----LMSESTSNIDQG-PRDHSVOL 1000
Db HDISI-----DYRFNHCKDHTVS---GDDEDYCPRSKANLGNASMTQHGTADEVAVET 981
QY 1001 PKPVHKPNRCFYSYSCQEQDLQILBALNSRGHRESALKETILOEKSRICAQLARFSE--EK 1058
Db TTPKQGNLWFLCDSOKELDELLNCHLPQIGRESQKLERL--EK-----RYODIHS 1031
QY 1059 FHFSDKPPQDSKPTYSRGRSSNAYDPSQWCAEQLELRDLRDLFDIEDRIYOGTLGAKV 1118
Db IHLARKPNLGLK-----SCDGNQ-----ELLNFRSLDIEVATRLQKGLGYVEE 1076
QY 1119 TDRHWRSALESGRYELSEENKENGIIKTVDNEVEEMEIDEQTKVIVK--DRLLAGIKTE 1176
Db T-----SEFEA---RVISLEKL-----KDFGECVIALQASVKFQLQGFMAPOK 1118
QY 1177 TPSTVSTWASTPOSVSSVHYLAMALFOIEQIERFLKAPLADSDSRSTYKTVLDWRRE 1236
Db RKLOQESAKTEEYDE-----EKQWEEAKVAS-----ALEKWK 1154
QY 1237 SLLSASLSOVFLHSLTLDRSVIMSKITLNARCKICRKGDAENNVLCGDCGRGHHTYCV 1296
Db AIREAQTSRMHVLLGLMDACIKWMSAENARCVCPKKGDDDKULCDECNKAPHLFCL 1214
QY 1297 RPKLTVPEGWFCPECPKORCRRLSPQRPSLESDESDVEDSMGDEDDVDGDEEQQS 1356
Db RPALYEPDGEWQCPACQATARRNSGRNYTEESASEDSEDESEEE-----EESEEE 1270
QY 1357 EEEVEVEQDEDDSOEEBEVSLPKRGPQVRL-PVKT-RGKLSSSFSGQQQEPGRYPS 1414
Db EEDYEV-----AGLRLPRKTIRGKHSVIPPAAARSGRRPGKPH 1310
QY 1415 RSQOSTPTVTSKTSRSLKINSAPPTETKSLRIASRSTRSHGPHLQADVPVELLSPPR 1474
Db STRRSQPK-----APPVD-----DAEVEL----- 1330
QY 1475 KRRGRKSANTPNSPNFNFRIATKSEQSRSNVNIASKLSLQESKRRCKRQSP 1534
Db -----VLQTKSSRRQSL----- 1343
QY 1535 SPVTLGRSSRGQGVHLSAFELQVLVLRHDDSWPFLKVLKSVTKQVPDYDIIKKPIL 1594
Db -----ELQCEETLHKIVKYRFSWPPREPVRDEAEDYDVITHPMDF 1386
QY 1595 NIIREKVNKCEYKLASEFTDIDELMFNSCFNVPENT 1631
Db QTVQNKCGSRYRVOEFLDMKQVFTNAEVNCRGS 1423
```

RESULT 12

AAW81171.

ID AAW81171 standard; protein; 1972 AA.

XX

```
AC AAW81171;
XX 05-MAY-1999 (first entry)
XX Human BAZ2-beta protein.
XX Transcriptional regulator; transcription; BAZ1-alpha; bromodomain; BAZ;
XX atypical zinc finger; testis; human; tumour; BAZ1-beta; BAZ2-alpha; drug;
XX BAZ2-beta; treatment; cancer; proliferative disorder; screening.
XX Homo sapiens.
XX WO9847920-A1.
XX 29-OCT-1998.
XX 17-APR-1998; 98WO-JP001783.
XX 18-APR-1997; 97JP-00116570.
XX 24-OCT-1997; 97JP-00310027.
XX (CHUS ) CHUGAI RES INST MOLECULAR MEDICINE INC.
XX Jones MH;
XX WPI; 1998-583603/49.
XX N-PSDB; AAV68403.
XX Transcriptional regulator gene family containing bromodomain - may be
XX expressed in testis tissue and is useful for treatment of cancer and
XX other proliferative disorders.
XX Claim 1; Page 100-116; 187pp; Japanese.
XX This sequence represents the human BAZ2-beta protein, a member of a
XX family of transcriptional regulator genes containing a bromodomain (BAZ,
XX Bromodomain with Atypical Zinc finger) which are expressed specifically
XX in testis tissue and also in certain tumour lines. Transgenic cells may
XX be used for the preparation of the BAZ1-alpha, BAZ1-beta, BAZ2-alpha and
XX BAZ2-beta proteins. These proteins can be used in the treatment of cancer
XX and other proliferative disorders, and in screening of compounds for
XX their binding ability to the expression products (e.g. for use as drugs
XX by modulation of transcriptional regulation)
XX
XX Sequence 1972 AA;
Query Match 8.4%; Score 727.5; DB 2; Length 1972;
Best Local Similarity 21.6%; Pred. No. 1.1e-40;
Matches 338; Conservative 190; Mismatches 427; Indels 613; Gaps 53;
QY 420 SPANRRRGPPKRIHISOEDNV-----DKEKKREELKIVVEERLKKKEERLKVREKE-- 527
Db 664 SRMRKGRPPVNVDAEFLDNADAKLRLKQAOBIAQAAQIKLRLKQEQEQAQVAKA 723
QY 443 NKQ--TLASVRSKATKRDKLLKQEE-----MKSIAFE-----KAKLKREKADA- 484
Db 724 KQQAIAAAEKKRQKEQIKIMQKEKTKRIQIRMEKELRAQOILEAKKKKEEAAAK 783
QY 485 -LEAKK--KEKE-----DKEKKREELKIVVEERLKKKEERLKVREKE-- 527
Db 784 LLEAKRIKEKEMRRQOAVLLKHQERRRRRHHMLMKAMEARKKAEKERLQKREKDR 843
QY 528 --REKLREKKYVEYLKQMSKPRDECDLKEP--PTPVKTRLPPEIFGDALMWLEF 584
Db 844 LNKERKLEQRLELEMAKELKKNEDMCLADQKLPPLPRIPGLVLSSTSDCLMVVQF 903
QY 585 LNAFGEL--FDLQDFPDGVTLEVLEBALVG-NDSEBGLPCLLFFFLTAIFAIAESEE 641
Db 904 LRNFQVLGFDVNDVFN---LSVLQEGLLNIGSMGEVQDLLVRLLSA---AVCDPLGI 957
QY 642 VAKELQTDATYKGCGLSKSLDLDSCITLSEILRLHLASGADVTSANAKYVQKRGFFDATD 701
Db 958 TGYRAKT---ALGHEHLLNVGNVNRDNVSEILQIF----- 987
```


PR 29-OCT-2002; 2002US-0422472P.
 XX (GETH) GENENTECH INC.
 XX Aggarwal S, Clark H, Gurney AL, Schoenfeld J, Williams PM;
 PI Wood WI, Wu TD;
 XX WPI; 2004-376182/35.
 DR N-PSDB; ADP54419.
 XX
 PT New PRO polynucleotides and polypeptides, useful in diagnosing
 PT and treating an immune related disease, e.g. systemic lupus
 PT erythematosus, rheumatoid arthritis, diabetes mellitus or asthma and in
 PT stimulating an immune response.
 XX
 PS Claim 1; SEQ ID NO 396; 3009pp; English.
 XX
 CC The present invention describes an isolated PRO nucleic acid (1). Also
 CC described: (1) a vector comprising (i); (2) a host cell comprising the
 CC vector of (1); (3) a process for producing a PRO polypeptide; (4) an
 CC isolated PRO polypeptide; (5) a chimeric molecule comprising the
 CC polypeptide of (4) fused to a heterologous amino acid sequence; (6) an
 CC antibody which specifically binds to a polypeptide of (4); (7) a
 CC composition of matter comprising a polypeptide of (4), an agonist or
 CC antagonist of the polypeptide or an antibody that binds to the
 CC polypeptide in combination with a carrier; (8) an article of manufacture
 CC comprising a container, a label on the container and a composition of
 CC matter of (7); (9) a method of treating an immune related disease in a
 CC mammal; (10) a method for determining the presence of a PRO polypeptide
 CC in a sample suspected of having the polypeptide; (11) a method of
 CC diagnosing an immune related disease or an inflammatory immune response
 CC in a mammal; (12) a method of identifying a compound that inhibits or
 CC mimics the activity of or expression of a gene encoding a PRO polypeptide
 CC ; and (13) a method of stimulating the immune response in a mammal. The
 CC PRO sequences have anti-allergic, anti-nausea, anti-arthritis,
 CC anti-rheumatic, anti-diabetic, anti-inflammatory, antipsoriatic,
 CC anti-hemostatic, hepatotropic, immunostimulant, immunosuppressive, muscular,
 CC haemostatic, neuroprotective, osteopathic, respiratory, vasotropic and
 CC virucide activities, and can be used in gene therapy. The nucleic acid
 CC (1) and the encoded polypeptides, compositions, kits and methods are
 CC useful in diagnosing and treating an immune related disease and in
 CC stimulating an immune response. The present sequence represents a human
 CC PRO protein from the present invention.
 XX Sequence 1972 AA;
 SQ
 Query Match 8.4%; Score 726.5; DB 8; Length 1972;
 Best Local Similarity 21.4%; Pred. No. 1.3e-40;
 Matches 335; Conservative 195; Mismatches 425; Indels 613; Gaps 53;
 QY 420 SPANRRGRPP-----KRIHSQDNVA--- 442
 DB 664 SMRRRRKRPNVGNAEPLDNADAKLLRLKLAQEIARQAQIKLLRLKQKQEQARVAKEA 723
 QY 443 --NKOTLASYSKATKEDDKLLKQEE-----MKSIAFE-----KAKLKEKADA- 484
 DB 724 KQQAQIMAAEBKRQKEQIKMKQOEKIKRIQIRMEKELRAQQILEAKKKKGEANAK 783
 QY 485 -LEAKK--KEKE-----DKEKKREBELKIVBEERLKKKEERLKVREKE-- 527
 DB 784 LLEAEKRIKEMRRQQAVALLKHOERERRRRHMLMKAEARKKAEERLKKQEKDEKR 843
 QY 528 --REKLREKKYVEYLKQNSKPRDEMCDDLKELPE-PTPVKTRLPPEIFGDALMVLEF 584
 DB 844 LNKERKLQRLELEMAKELAKPNEDMCLADQKPLPELPRIPLGLVLSGSTFSDCLMVVQF 903
 QY 585 LNAFGL--FDLQDFPFGVTLVLEALVG-NSEGPLCELLPFFLTAIQAABEE 641
 DB 904 LRNFGLVGFVNIDVFN--LSVLQEGLLNIGSMGEVQDILLVRLISA---AVCDPLGI 957
 QY 642 VAKQLTDADTKGCSLKLSDLSCTLTSEILRLHILASGADVTSANAKRYQKRGGFATD 701

DB 958 TGYKAKT---ALGEHLLNVGVNRDNVSEILQIF----- 987
 QY 702 DACMELRLSNPSLVKKLSSTSVYDLTPGEKMKILHALCGKLL---TLVSTRDFIEDYVDI 758
 DB 988 ---NEAHCQGOTELTESLTKAFQAHTPAQKASVLAFLINELACSKSVSEIDKNIDYNSN 1044
 QY 759 LRQAKQBFRELKAEQHRKEREAAARIRKKEEKLKEQEQMKKEQKQKLEKDEQNSTAD 818
 DB 1045 LRDKWVV-----EGKLR-----KLRIIHAKTGRDTSGG 1075
 QY 819 ISIGEEE-----REDPDTIESKDTQKELDQDMFTEDDDPQSHKRGRRGK 865
 DB 1076 IDLGEQHPGLTPTPGRRKRRKGGDSYDDDDDDSDQDQDEDEDEED---KEDQKKG 1131
 QY 866 RGQNGFKFPTREQINCVTRELTLTADREE---ALKQEHQKKEKELLEKIOS-----A 914
 DB 1132 KTD-----ICEDDEGQAAASVEELEKQIEKLSKQSQYRRKUFDA 1172
 QY 915 IACTNIFPLGRDRMYRWYIPPSIPGLFIE--EDYSGLTEDMLLPRPSSFQNNVQSOPQ 972
 DB 1173 SHSLRSVMFGPDRYRRYRWWILPRCGGIFVEGMESGEGLEELAKEREKKAESVQIKEEM 1232
 QY 973 VSTKTGPLMSESTNIDQG-----PRHVSOLP 1001
 DB 1233 FET-SGDSLNCNTHCEQKEDLKEKDNLTFLQKPGSFSKLSKLELVAKMPPSEVMTP 1291
 QY 1002 KP-----VHK-----PNRWCFYSCEQ 1018
 DB 1292 KPNAGANGCTLSYQNSGKHSLSGVSQATOSNVEKADSNLNFNTGSSGPGK--FYSPLPN 1349
 QY 1019 LDQILTEALNRGHRRESAL-----KETLLOEKSRICQALARFSEKHFSDKPOPD--- 1068
 DB 1350 -DQLLKLTENRQWFSLLPRTPCDDTSLTHADMTASLVTVPQSPFSKSPSPAPLGS 1408
 QY 1069 -----SKPTYSRG-----RS 1078
 DB 1409 SAQNPVGLNPFALSPLOVKGVSMMGLQFCGWPTGVVTSNIPFTLSVPSLGSGLSGBN 1468
 QY 1079 SNAYDPSQCAEQ-----LEL-RLRDF-----LLDI 1104
 DB 1469 GNSFLTNSVASSKESVPPQNEKATSAQAAVEAVAKPVDPPSPKPIPEEMQFGWRIIDP 1528
 QY 1105 EDRIYQGTLAGI-----KVTDH---IWRSALESGRYELLIS--EENKENG 1145
 DB 1529 ED-----LKALLKVLHRLGIREKALQKQKHLDTYITQACLKNKDVAIIELNENEQV 1582
 QY 1146 IKTV--NEDVEEMEIDQTKVIVKDRLLGIKTETPTSTVSTNASTPQSVS----- 1192
 DB 1583 TRDIVENWSVEEQAMENDLSVLQQVEDLERRVASASLQVKGWMCPEPASEREDLVYFEHK 1642
 QY 1193 -----SVVHYL-----AMALFQIEQGIERRFLK--APLDAS 1221
 DB 1643 SFTKLCKEHDGEFTGEDSSAHLERKSDNPLDIATVTRLADLERNIERRIEDIAP---- 1698
 QY 1222 DSGRSYKTVLDRWRRESLLSSASLSQVFLHSLTLDRSVIWSKSIILNARCKIKRKGDAENM 1281
 DB 1699 -----GLRVRRARLSEARSAQAVALCTQLOKQSTAEWSKIMKYVCQICRKGDNELL 1750
 QY 1282 VLCDGCDRGHHYTCVRPKLKTVPBGDFWFCPECRKQ-----RCRRLSFRQPSLESBDV 1336
 DB 1751 LLCDCGDKGCHTYCHRPKITTIIPDGWFCPACIAKASQQLIKKLVHVKKTTNESKK-- 1808
 QY 1337 EDSMGGEDDEVDGEESQSEEEYEVEQEDDDSQEESVSLPKRGPPQVPLPVKTRGKL 1396
 DB 1809 -----GRKVTLTGDE-----DED----- 1822
 QY 1397 SSSFSRQSQQPCRYPSRQSQSTPKTTVSSKTSGRSLRKINSAPPTETKSURIASRSTRH 1456
 DB 1823 -----SASTSSSL----- 1830
 QY 1457 SHGPLQADVFWELLSPRRKRGRRKSANNTPENSPFNFRVIATKSSEQSRSVNIASKLS 1516
 DB 1831 -----KRGKND-----LQKRKMEENTSIN---LS 1851


```
Db 1099 PRGIREKALHKLHNRDFOEVCILRPSADPIF-----EPRQLPAFOEGIMS--WSPKEK 1151
Qy 1088 CAEKQLELRDLRDLIDEDRYVQGLGAIKVTDRHIMESALESGHYELLSEENKENGIIK 1147
Db 1152 TYE--TDLAVLQWBEELSQRVIM-----SDIQIRGWTCPSPD----- 1186
Qy 1148 TVNEDVEEME--IDQTKVIVKDRLLGIKTETPSTVNASTPQSVSVVHYLAMALFQI 1205
Db 1187 STREDLAYCEHLSQEDITWRGR--GREGLAPQRKTTNP-----LDLAVRLAAL 1235
Qy 1206 EOGIERRELFKAPLSDASGRSVKTVLDR-----WRESL 1238
Db 1236 EQNVERRYLRREPL-----WPTHEVLEKALLSTPNGAPEGTTTETSYEITPRIRVWRQTL 1290
Qy 1239 LSSASLSOVFLHLSLTLDSVYWSKILNARCKICRKGDAENWVLCDCDGHHTYCVRP 1298
Db 1291 ERCSAAQVCLCGQLERSIAWEKSVNWKVCLVCRKGDNDFFLLCDGCDRGCHIYCHRP 1350
Qy 1299 KLKTVPEGDFWCFECPKRCRRLSFRQPSLESDEDDVEDSMGGDDDDVDGDEEGQSEE 1358
Db 1351 KWEAVPEGDFWCTVCLAQQ-----VEGEFTQKFPFPRGQXK 1388
Qy 1359 BEYVEQEDDSQEEVEVSLPKRGRQVRLPVKTRGKLSFSRSGQOQOQEPYPSRSQ 1418
Db 1389 SGYSLNPFSEG-----GR--RRVLLRGR-----ESPAAGPRYSEE 1422
Qy 1419 STPKTVTSKGRSLRKINSAPPTETKSLRTASRSTRSHGFLQADVVELLSPRKRG 1478
Db 1423 G-----LSPSKRRR- 1431
Qy 1479 RKSANNTPENGPNPNFRVIATKSEQSRSVNIASKLSLOESKRRCKRQSPSPSPVT 1538
Db 1432 -----LSMRNHS----- 1439
Qy 1539 LGRSSRGQGVHLSAEQLVVELVRHDDSWPFLKLVSKIQVPDYDIIKKPIALNIIR 1598
Db 1440 -----DLTFCEIILMEMESHDAAPFLEPNPRLVSGYRIIKNPMDFTMR 1486
Qy 1599 EKVNKCEYKLASEFIDDELMSFCNFYFNPRNTSEAKAGTLQAPF 1644
Db 1487 ERLLRGGYTSSEFAADALLVDFNQCFTNEDDSEVGKAGHMRRPF 1532

RESULT 16
ID AAB95554 standard; protein; 1727 AA.
AC AAB95554;
XX
XX
DT 26-JUN-2001 (first entry)
XX
DE Human protein sequence SEQ ID NO:18183.
XX
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX Homo sapiens.
XX EP1074617-A2.
XX
XX 07-FEB-2001.
XX
XX 28-JUL-2000; 2000EP-00116126.
XX
XX 29-JUL-1999; 99JP-00248036.
PR 27-AUG-1999; 99JP-00300253.
PR 11-JAN-2000; 2000JP-00118776.
PR 02-MAY-2000; 2000JP-00183767.
PR 09-JUN-2000; 2000JP-00241899.
XX
XX (HELI-) HELIX RES INST.
XX
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
```

```
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI; 2001-318749/34.
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602 full-
XX length cDNAs defined in the specification, and for the detection and/or
XX diagnosis of the abnormality of the proteins encoded by the full-length
XX cDNAs.
XX
XX Claim 8; SEQ ID NO 18183; 2537pp + Sequence Listing; English.
XX
XX The present invention describes primer sets for synthesizing 5602 full-
XX length cDNAs defined in the specification. Where a primer set comprises:
XX (a) an oligo-dT primer and an oligonucleotide complementary to the
XX complementary strand of a polynucleotide which comprises one of the 5602
XX nucleotide sequences defined in the specification, where the
XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination
XX of an oligonucleotide comprising a sequence complementary to the
XX complementary strand of a polynucleotide which comprises a 5'-end
XX sequence and an oligonucleotide comprising a sequence complementary to a
XX polynucleotide which comprises a 3'-end sequence, where the
XX oligonucleotide comprises at least 15 nucleotides and the combination of
XX the 5'-end sequence/3'-end sequence is selected from those defined in the
XX specification. The primer sets can be used in antisense therapy and in
XX gene therapy. The primers are useful for synthesizing polynucleotides,
XX particularly full-length cDNAs. The primers are also useful for the
XX detection and/or diagnosis of the abnormality of the proteins encoded by
XX the full-length cDNAs. The primers allow obtaining of the full-length
XX cDNAs easily without any specialised methods. AAH03186 to AAH13628 and
XX AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893
XX represent human amino acid sequences; and AAH13629 to AAH13632 represent
XX oligonucleotides, all of which are used in the exemplification of the
XX present invention
XX
XX Sequence 1727 AA;
XX
XX Query Match 8.0%; Score 689.5; DB 4; Length 1727;
XX Best Local Similarity 20.3%; Pred. No. 4.1e-38;
XX Matches 314; Conservative 195; Mismatches 387; Indels 649; Gaps 48;
Qy 424 RRRGRPPK-RIH--ISOEDN-----VANKQTIASYRSKATKRDLLK---QDEM 467
Db 493 RGRGRPPKVKITELLNKTNRPLKLEAQETLNEEDKAKIAKSKKQKQVQEGCOTTI 552
Qy 468 KSLAFKAKLREKADALEAKKKEKEDK---KREELKKIVEE-----RLKKKEE--- 516
Db 553 QGQARNRKRQETKSLKQKAEKSKKABKGBKTKQEKLEKVKREKKEKVKQKKEBVTK 612
Qy 517 -KERLKVREKEREREKREKRYVEYLKQWSKPREDMCEDDLKELPBPVPVK--TRLPEI 574
Db 613 AKPACKADKTLATQRLREERQROQMILEEMKKPTEDMCLTDHQLPDSRVPGTLPSGA 672
Qy 575 FGDALMVLEFLNAPGEL--FDLQDEFPDGVTVLELEALV-GNDSEGLPCELLFFFTAI 631
Db 673 FSDCLCTIVEFLHSPKVLGFDPAKDV---SLGVLEGLLCCQDGLSEVQDVLRLKAA 729
Qy 632 F-----QAIAEEEEVAKEQLTDADTKGCSKSLDLDSCITSEILRLHILASGADV 682
Db 730 LHDGFPSPYCSQSLKILGEKVEIPLTRDN-----VSEILRCFLMAYGVE- 773
Qy 683 TSANAKRYQKRGGFDDATDDACMELRLSNPSLVKLSSTSVYDLTPGKKIL---HAL 738
Db 774 -----PALCDRLTQFPQAPPOQAALVLAFLVHEL 804
Qy 739 CGKLLTLVSTRDFTIEDYVILRQAKQEFRELK--AEQHRKEREEBAAAIRKRBEKLEKQ 796
Db 805 NGSTL-----LINEIDKTLSEMSYRKNKKNVIGERLRLKTVLAKRTGRSEVEMG 856
Qy 797 EQKM-KEKQEKLEDEBORNSTADISIGEEEREDFTSIESKDTQEKELDQDMFTDEDDP 855
Db 857 EECLEGRSSRIIMEE-----TSGMEE-----EEEES 883
Qy 856 GSHKRRGRGKRG-----QNGKFTFQEQINCVTRELLTADDEEALKQEHORKEKELLE 909
```


Db	884	IAAVPGRGRGRDGEVDATASSIPELERQ-----IEKLSKRQLFRKKLLH	928
Qy	910	KIQSAIACTNIFPLGRDMYRRYWIFFSIPGLFIEEDYSGLTEDMLP-----	957
Db	929	SSQMLRAVS-----LGQDYYRRYYVWLYLAGIFVEG-----TEGTVPEEVIKETDSLK	979
Qy	958	-----RPSSFQ-----NNVQSQD-----	970
Db	980	VAHASLNPALFSMKMELAGSNTTASSPARARGPRKTKPGSMOPRHLSKPVRGQDSEQ	1039
Qy	971	-----POV-----STK	976
Db	1040	QAQLQPAQAHPAQPOQLOLQLOSHKGFLEQSGSLSLGQSOHDLGQSAFSLWSQTO	1099
Qy	977	TGEPLMESTSNIDQGR-----DHSVQLP-----	1001
Db	1100	SHSLLSSSVLTPDSSFGKLDPAFQPPPEPDEASSPDQALWENISQAHPCAAPT	1159
Qy	1002	-----KPVHKPNR-----	1009
Db	1160	PPPAVSDQTPSPQOLASSKPMNRPAAANPCSPVQFSTPLAGLAPRRAGDPGEMPOS	1219
Qy	1010	-----WCFYSSCEOLDQLEIALNS	1028
Db	1220	PTGLGQPKRRGRPPSKFFKQWEQRYLQLTAAQVPPEMCGSMWIRPEMLDAMLKALHP	1279
Qy	1029	RGHRESALKETLLQEKESRIQAULARFSEEEKFHFSKDPQDSKPTYSGRSSNAYDPSQMC	1088
Db	1280	RGIREKALHKLHNKGRDFLOEVCLRPADPIF-----EPQLPAPQEGINS--WSPKEKT	1332
Qy	1089	AEKOLELRLRDFLDIEDRIYQGTGLGAIKVTDRIHWSALESGRGYELLSEENKENGIIKT	1148
Db	1333	YE--TDLAVLQWBELEQRVIM-----SDLIQRGWTCPSPD-----S	1367
Qy	1149	VNEDVEEME--IDBQTKVIVKDRLLGIKTTPTSTVSTNASTPQSVSSVWHYLAALFQIE	1206
Db	1368	TREDLAYCEHLSQSDITWRGR--GREGLAPORKTTNP-----LDLAVNRLAAL	1416
Qy	1207	QGIERRELFKAPLDASDSGRSYKTVLDR-----WRESLL	1239
Db	1417	QNVERRYLRPL-----WPTHEVVLKALLSTPGAPEGITTEISYEITPRIVRWQTLE	1471
Qy	1240	SSASLSQVFLHSLTLDRSVITWSKILNARKICRKGDAENMVLCDGCDRGHTTYCVRPK	1299
Db	1472	RCRSAAQVCLCLGOLERSIAWEKSVNKTCLVCRKGDNDFFLLLCDGCDRGCHYCHRPK	1531
Qy	1300	LKTVPEDGWCPCEPRKQRCRRLSFRQPSLESDEVDSDNGCEDDEVGDDEEGQSEEE	1359
Db	1532	MEAVPEGDWCTVCIAQQ-----VEGEFTQKPGFKRGQKRKS	1569
Qy	1360	EYEVQEDDSDSQUEEEVSLPKRGRPOVRLPVKTRGKLSGSFSGRQOQEPGRYPSRQQS	1419
Db	1570	GYSLNFSEGD-----GR--RRVLLRGR-----ESPAAGPRVSEEG	1603
Qy	1420	TPKTTVSGKTRSLRKINKINSAPPTETKSLRIASRSTRHSHGLOADVVELLSPRKRRGR	1479
Db	1604	-----LSPSKRRR-----	1611
Qy	1480	KSANNTPENSPNFPNFRVIATKSSQESRSVNIASKLSQESKRCRCKQSPSPVTL	1539
Db	1612	-----LSMRNHS-----	1619
Qy	1540	GRSSGRGGVHELSAFEQLVVELVRHDDSWPFLKVLVSKIQVPDYDIIKKPIALNIIRE	1599
Db	1620	-----DLTFCEIILMEMESHDAWPFLFPVNPRLVSGYRRIKNPMDFSWMRE	1667
Qy	1600	KVNKEYKLASEFTIDDLFMFSCNCFEYNPNRTSEAKGTLQAPF	1644
Db	1668	RLLRGGTYSSEFAADALLVFDNCOTFNEDDSEVGKAGHTWRAPF	1712

RESULT 17

AA040239	AA040239 standard; protein; 1878 AA.
ID	AA040239 standard; protein; 1878 AA.
XX	AA040239;
AC	AA040239;
XX	AA040239;
XX	AA040239;
DT	22-OCT-2001 (first entry)
XX	22-OCT-2001 (first entry)
XX	Human polypeptide SEQ ID NO 3384.
XX	Human; neurotropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW	peripheral nervous system; neuropathy; central nervous system; CNS;
KW	Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW	amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW	chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW	leukaemia.
XX	leukaemia.
OS	Homo sapiens.
XX	Homo sapiens.
XX	XX
PN	W0200153312-A1.
XX	W0200153312-A1.
XX	26-JUL-2001.
PD	26-JUL-2001.
XX	26-DEC-2000; 2000WO-US034263.
PF	26-DEC-2000; 2000WO-US034263.
XX	26-DEC-2000; 2000WO-US034263.
PR	23-DEC-1999; 99US-00471275.
PR	21-JAN-2000; 2000US-00488725.
PR	25-APR-2000; 2000US-00552317.
PR	20-JUN-2000; 2000US-00598042.
PR	19-JUL-2000; 2000US-00620312.
PR	03-AUG-2000; 2000US-00653450.
PR	14-SEP-2000; 2000US-00662191.
PR	19-OCT-2000; 2000US-00693036.
PR	29-NOV-2000; 2000US-00727344.
XX	29-NOV-2000; 2000US-00727344.
XX	(HYSE-) HYSEQ INC.
PA	(HYSE-) HYSEQ INC.
XX	(HYSE-) HYSEQ INC.
PI	Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI	Pang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;
PI	Zhou P, Goodrich R, Drmanac RT;
XX	Zhou P, Goodrich R, Drmanac RT;
XX	WPI; 2001-442253/47.
DR	N-PSDB; AA159395.
DR	N-PSDB; AA159395.
XX	N-PSDB; AA159395.
XX	N-PSDB; AA159395.
PT	Novel nucleic acids and polypeptides, useful for treating disorders such
PT	as central nervous system injuries.
XX	Novel nucleic acids and polypeptides, useful for treating disorders such
PS	as central nervous system injuries.
XX	as central nervous system injuries.
PS	Example 5; SEQ ID NO 3384; 10078pp; English.
XX	Example 5; SEQ ID NO 3384; 10078pp; English.
XX	Example 5; SEQ ID NO 3384; 10078pp; English.
CC	The invention relates to human nucleic acids (AA157798-AA161369) and the
CC	encoded polypeptides (AA038642-AA042213) with neurotropic,
CC	immunosuppressant and cytostatic activity. The polynucleotides are useful
CC	in gene therapy. A composition containing a polypeptide or polynucleotide
CC	of the invention may be used to treat diseases of the peripheral nervous
CC	system, such as peripheral nervous injuries, peripheral neuropathy and
CC	localised neuropathies and central nervous system diseases, such as
CC	Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC	lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC	utilisation of the activities such as: Immune system suppression,
CC	Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC	and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC	assays for receptor activity, arthritis and inflammation, leukaemias and
CC	C.N.S disorders. Note: The sequence data for this patent did not form
CC	part of the printed specification
CC	part of the printed specification

[illegible]

CC carcinoma including lymphoma. The present sequence is one such CAP. Note:
 CC This patent is an equivalent to basic patent US2002182586A1, for which no
 CC sequence data was published

XX
 SQ Sequence 1873 AA;

```

Query Match      8.0%; Score 688.5; DB 7; Length 1873;
Best Local Similarity 20.1%; Pred. No. 5.4e-38;
Matches 312; Conservative 198; Mismatches 378; Indels 663; Gaps 47;

QY 424 RRRGPPKRIHSQEDNVANKOTLASYSKATKRDKLLKQEMSKLAFKAKLREKAD 483
DB 641 RGRGPPK-VKTELLNKTDRPLKLEAQT-----LNEEDKAKIAKSKKM-RQKVQ 692
QY 484 ALEAK-----KKEKEDKKEKBEELKVIIEERLKKKEKELKVERE-KEREKLR- 532
DB 693 RGECQTIIQOARNRKQETSLKQEAKKSKKEKGTQKELKEKVKREKKEKVMKE 752
QY 533 -----EKKRYVEYLKQWSKPREMECDLKLPEPTPVK-T 568
DB 753 KEEVTAKPACAKDTLATQRLEERQOQMTLEEMKPTEDMCLTDHQLPDFSRVPL 812
QY 569 RLPPFIQDALMVLNFAFGL--FDLQDEPPDGVTLEVLLEALV-GNDSGGLCELLF 625
DB 813 TLPSCAFSDCLTIVFLHSFGVLGFDPAKVP---SLGVQEGLCQGDLSGEVQDILLV 869
QY 626 PFLTAIF-----QAIAEEEEVEAKEQLTDADTKGCSLKLSDLSCTSEILRLHIL 676
DB 870 RLLKAALHDPGPPSYQSLKILGKVSIEPLTRDN-----VSEILRCFLM 914
QY 677 ASGADVTSANAKRYQKGGGPDATDDACHLRLSNPSLVKLSVTSVVDLTPGEKML- 735
DB 915 AYGV- -----PALCDRLRTQPPQAQPPQKAAVLA 944
QY 736 ---HALCGKLLTVSTDFIEDYVILRQAKQEFRELK--AEQHKEREAAAARLKEKE 790
DB 945 FLVHNLNSTL-----IINEIDKTLSEMSYRNKWIIEGRULRLKTVLAKTGRSE 996
QY 791 EKLKEQEQKM-KEKQEKLEDEQRNSTADISIGEEEREDFDTSIESKDEQKELQDMFT 849
DB 997 VEMEGPECLGRRRRSIMEE-----TSGMEE----- 1023
QY 850 EDEDDPGHKRRGRKRG-----QNGFKETFRQEQINCVTRELLTADAEBAALQEHQK 903
DB 1024 EEEESIAAEPGRGRGRDGEVDATASSIPELERO-----IEKLSRQLFF 1068
QY 904 EKELLEKIQSACTNIPFLGRDRMYRYWIPPSIPGLFIEDYSGLTDMLLP----- 957
DB 1069 RKLLHSSQMLRAVS-----LGQDRYRRYVLPYLAGIFVEG-----TEGNLVPEEVIKK 1119
QY 958 -----RPSSEFQ-----NNVQS 968
DB 1120 ETDLSLKVAAHASLNPAFLSMKELAGSNNTASSPARARGRPRTKPGSMQPHLKSPPRG 1179
QY 969 QD-----POV----- 973
DB 1180 QDSEQOQAQLOPEAQLHAPAPQPOLQLOSHKGFLEQSGPLSLGQSQHDLSQAFLS 1239
QY 974 ---STKTGEPLWSESTNIDQGR-----DHSVQLP 1001
DB 1240 WLSQTQSHSLSSSVLTFDSSPGKLDPAQPPPEPEPEDEABSPDQALWFWNISAQMP 1299
QY 1002 -----KPVHKPNR----- 1009
DB 1300 CNAAPTPPVAVEDQTPSPQOLASSKPMNRPANPCSPVQFSSTPLAGLAPKRRAGDP 1359
QY 1010 -----WCFYSSCEQLDOL 1022
DB 1360 GEMPOSTGLGQPKRRGRPPSKFFKQMEQRVLTQLTAQVPPEMCSGWWIWRDPEMLDAM 1419
QY 1023 IEALNSRGHRESALKETILLOEKSRICAQLARPESEKHFPSDKPQDPDSKPTYSRGSSNAY 1082
DB 1420 LKALHPRGIREKALHKLHNRDFLQVCLRPSADPIF-----EPROLPAFQEGIMS--W 1472

```

```

QY 1083 DPSQCAEQLELRDLRDLFLDIEDRIYQGTGLGAIKVTDRIHWSALESGRYELLSEENKE 1142
DB 1473 SPKEKTYE--TDLAVLQWVEELEQRVIM-----SDLOIRGWTGCPSPD--- 1512
QY 1143 NGIITVNEDEME--IDEQTKVIVKDRLLIGIKTETSTVSTNASTQSVSSVVHYLAM 1200
DB 1513 -----STREDLAYCEHLSQSDIEDITWRGR--GREGLAPQKRTNP-----LDLAVM 1556
QY 1201 ALFOIEQIEERRFLKAPLDASDSGRSYKTVLDR----- 1233
DB 1557 RLAALEQNVERRYLYREPL-----WPTHEVVLKALLSTPNGAPEGTTEISYEITPRIV 1611
QY 1234 WRESLLSASISQVFLHLSTLDRSVIWSKSLNARCKI CRKKGDAENMVLCDGCDGRHHT 1293
DB 1612 WRQTLERCRSAAQVCLCLGQLERSIAWEKSVNKTCLVCRKGDNDÉFLLCLDCDGRGCHI 1671
QY 1294 YCVRPKLTVPBGDWFCPCBPQBCRLRSFRQPSLESDEDESDMGEDDEVGDDEE 1353
DB 1672 YCHRPQMEAVPEGDWFTVCLAQ-----VEGEFTQKGFPPKR 1709
QY 1354 GQSEEEYEVEDDEDSQEEEEVSLPKRGRPOVRLPVKTRGKLSSSFSSRGOQOQSPGRYP 1413
DB 1710 QQKRSYSLNPFSEGD-----GR--RRVLLRGR-----ESPAAGP 1743
QY 1414 SRQOQSTPKTTVSSKTGRSLRKINSAPPTETKSLRIASRSTRSHGHPQLQADVVELLSPR 1473
DB 1744 RYSEEG-----LSPS 1753
QY 1474 RKRGRKSNANTPENSFPNFRVIATKSSQSRSVNTASKLSLOESSEKRCRKRQSP 1533
DB 1754 KRRR-----LSMENHHS----- 1765
QY 1534 PSPVTLGRRSSRGQGVHELSAFQVLVVELVRHDDSWPFLKLVSKIQVDPYVDIHKPIA 1593
DB 1766 -----DLTFCEIILMEMESHDAAPFLBPVNPRLVSGYTRRIINPMD 1807
QY 1594 LNIIREKVNKCEYKLASEFIDDIBLMSFCNFYNPNTSEAKAGTRLOAFF 1644
DB 1808 FSTMRRLLRGYTSSEBEAADALLVFDNCQTFNEDDSEVGKAGHIMRRFP 1858

RESULT 19
AAW81170
ID AAW81170 standard; protein; 1878 AA.
AC AAW81170;
XX
XX 05-MAY-1999 (first entry)
XX Human BAZ2-alpha protein.
DE
XX Transcriptional regulator; transcription; BAZ1-alpha; bromodomain; BAZ;
KW atypical zinc finger; testis; human; tumour; BAZ1-beta; BAZ2-alpha; drug;
KW BAZ2-beta; treatment; cancer; proliferative disorder; screening.
XX Homo sapiens.
XX WO9847920-A1.
PN
XX 29-OCT-1998.
PD
XX
XX 17-APR-1998; 98WO-JP001783.
PF
XX 18-APR-1997; 97JP-00116570.
PR
XX 24-OCT-1997; 97JP-00310027.
XX
XX (CHUS ) CHUGAI RES INST MOLECULAR MEDICINE INC.
PI Jones MH;
XX
XX WPI; 1998-583603/49.
DR N-PSDB; AAV68396.

```

```

XX Transcriptional regulator gene family containing bromodomain - may be
PT expressed in testis tissue and is useful for treatment of cancer and
PT other proliferative disorders.
XX
PS Claim 1; Page 72-88; 187pp; Japanese.
XX
CC This sequence represents the human BAZ2-alpha protein, a member of a
CC family of transcriptional regulator genes containing a bromodomain (BAZ,
CC Bromodomain with Atypical Zinc finger) which are expressed specifically
CC in testis tissue and also in certain tumour lines. Transgenic cells may
CC be used for the preparation of the BAZ1-alpha, BAZ2-alpha and
CC BAZ2-beta proteins. These proteins can be used in the treatment of cancer
CC and other proliferative disorders, and in screening of compounds for
CC their binding ability to the expression products (e.g. for use as drugs
CC by modulation of transcriptional regulation)
XX
SQ Sequence 1878 AA;

Query Match          7.9%; Score 681.5; DB 2; Length 1878;
Best Local Similarity 19.9%; Pred. No. 1.7e-37;
Matches 306; Conservative 198; Mismatches 397; Indels 639; Gaps 44;

QY 424 RRRGPPK-RIH--ISQEDN-----VANKOTLASYSKATKRDKLLKOEKMSL 470
DB 644 RRGPPKVKITELLNKTDNRLPKLEAQETFLNEDKAKIAKSKKQKQVQRECLTTI 703
QY 471 AFEKAKLKREKADAL---EAKKEKEDKEK---KREELKKTIVBEE-----RLKKKE--- 516
DB 704 QQARNRKQETKSLKHKEAKKSKAEKKGTKQEKLEKVKREKKEKVKMKSEEVTK 763
QY 517 -KERLKVEREKRLREKRYEYLKQWKPREDMECDLDELPEPTPVK-TRLPEEI 574
DB 764 AKPACKADKTLATORRLEERQKQMILEMKKPTEDMCLTDHQLPDRFVRPGTLPSGA 823
QY 575 FGDAWLVEFLNAGEL--FDLQDEFDPGVLEVEALV-GNDSEGLPCLELFFFLTAI 631
DB 824 FSDCLTIVEFLHSFGKVLGFDPADKP--SLGVLOEGLLCQGSLSGEVQDLVRLXAA 880
QY 632 F-----QAIAEBEEBAKQTDADTKGCSLSKSLDLSCTLSEILRLHIILASGADV 682
DB 881 LHDPGFSYCQSLKILGEKVSEIPLTRDN-----VSEILRCLFLMAYGV-- 923
QY 683 TSAKAKYIKRGGFDDTDDACMELRLNPSLVKLSSTSVYDLTPGCKMIL-----HAL 738
DB 924 -----KPALCDRLRTQFPQAPPOQKAAVLAFFPVHEL 955
QY 739 CGKLLTLVTRDFTDYDILRQAKQEPRELKAEQREEREAARIRKKEKLEKQEQ 798
DB 956 NGSTL-----INEIDKTLSEMS--YRKNKIVEGRLRLKTVLAK----- 995
QY 799 KMKKEQKLEDEQRNSTADISIGBEEREDFTSIESKDTQKQLDQDMFTEDDDPGSH 858
DB 996 -----RTGRSEVMGRPE---ECLGRSSSRIMETSQWEEEEEESTAA 1037
QY 859 KGRGKKG-----QNGFKFTFQEQINCVTRELLTADDEEALKQEHQKKEKLEKIQ 912
DB 1038 VPGRRGRDGEVDATASSIPELQ-----IEKLSKRLQFLFRKLLHSSQ 1082
QY 913 SAIACTNIFPLGRDMRYRRYIFPSIPGLFTIEDYSGLTEDMLLP----- 957
DB 1083 MLRAVS-----LQDQYRRRYVLPYLAGIFVEG-----TEGNLVPVEVVKETDLSLKVA 1133
QY 958 -----PPSSFQNNVQSQD----- 970
DB 1134 HASLNPAFLSMKMWELAGSNNTTASSPARARSRLTKPGFMQPRHFKSPVRGDSQPQQA 1193
QY 971 -----PQV-----STKGE 979
DB 1194 LQPEAQLHVPAPQPOLQLOLQSHKGFLEQESPLSQSHDLQSASFSLWLSQTSQHS 1253
QY 980 PLMSSTSNIDGPR-----DHSVQLP----- 1001

```

```

DB 1254 SLSSSVLTDPDSGKLDPAFPQPEEPPEDEABESSPDLOAFWENISQAQPCNAAPTPPL 1313
QY 1002 -----KPVHKPNR----- 1009
DB 1314 AVSEDOPTSPQQLASSKPMNRPSAANPCSPVQFSSTPLAGLAPKRAGDGENPQSPGTG 1373
QY 1010 -----WCFYSSCEQLDOLIEALNSRGH 1031
DB 1374 LQQPKRRGRPPSKFFKQMEQRYLTQLTQAQVPPPEMCSGWWIPPEMDLMDLKAHPRGI 1433
QY 1032 RESALKETLLQEKSRICAQLARFSEKPHFSDKPPQDSKPTYSRGRSSNAYDPSQMCABK 1091
DB 1434 REKALHKLHNRDFLQEVCLRPSADPIF-----EPRLDPAFQSGIMS--WSPKEKTYE- 1485
QY 1092 QLELRLEDPLLDIEDRIYQGTGLGAIKVTDRIHWSALESGRYELLSENKENGIIKTWNE 1151
DB 1486 -TDLAVLQWVELEQRVIM-----SDLQIRGWTCPSPD-----STRE 1521
QY 1152 DVEEMEIDEOTKVIKDRLLGIKTETPTSTVNASTPQSVSVVHYLAMALFOIEQOIER 1211
DB 1522 DLAYCEHLSQEDITWRGPGREGLAQKTTNP-----LDLAVMLAALEQONVKR 1572
QY 1212 RFLKAPLDASDSGRSYKTVLDR-----WRESLSSASL 1244
DB 1573 RYLREPL-----WPTHEVLEKALLSTPNGAPEGTTTTEISYEITPRIRIWRQTLQRCRSA 1627
QY 1245 SQVPLHLSTLDRSVIWSKILNARCKICRKGKDLENWVLCDGCDRGHTTYCVRLKLTVP 1304
DB 1628 AHVCLCLGHILERSIAWEKSNVKTCLVCRKGDNDDEFLLLCDGCDRGCHYIYCHRPKMEAVP 1687
QY 1305 EGMDFCEPCPKORCRRLSRFRQPSLESDEVEDSMGDEDDVDGDEEGQSEEEVEVE 1364
DB 1688 EGMDFCTVCLAQQ-----VEGEFTQKGFPGKQKRGKSGYSLN 1725
QY 1365 QDEDSQEEEEVSLPGRGPQVRLPVKTRGKLSSSFSSRGOQBPGRYPSSQOOSTPKTT 1424
DB 1726 FSEGD-----GR-----RRRVLLKGR-----ESPAAGPRYSE----- 1752
QY 1425 VSKTGSLRKLNSAPPTETYSKLIASRSTRSHGFLQADVVELLSPRRKRGKRSANN 1484
DB 1753 -----ERLSPSKRR----- 1762
QY 1485 TPENSPNPNFRVIATKSSQSRSVNIASKLQSESKRCKRQSPSPVTLGRSS 1544
DB 1763 -----LSMRNHS----- 1770
QY 1545 GROGVHLSAFEQVLVVRHDDSWPFLKLVSKIQVPDYDIIKKPIALNIIEKNKC 1604
DB 1771 -----DLTFCIILMEMESHDAWFFIEPVNRLVSGYRRIIKNPMDFSTMERLLRG 1823
QY 1605 EYKLASEFIDDIELMPSNCFEYNPNRTSEAKAGTLOAFF 1644
DB 1824 GYTSSEFAADALVFDNCQTFNEDDSEVGKAGHIMRRFF 1863

RESULT 20
ABM85416
ID ABM85416 standard; protein; 1586 AA.
XX
AC ABM85416;
XX
DT 18-NOV-2004 (first entry)
XX
DE Mouse protein sequence MCP4389.
XX
KW Cytostatic; carcinoma; lymphoma; cancer; murine.
XX
OS Mus musculus.
XX
FN WO2003073826-A2.
XX
PD 12-SEP-2003.
XX

```


arthritis; malignancy; digestive; endocrine; infection.

OS Homo sapiens.

PN WO9918208-A1.

XX 15-APR-1999.

XX 01-OCT-1998; 98WO-US020775.

XX 02-OCT-1997; 97US-0060833P.

XX 02-OCT-1997; 97US-0060836P.

XX 02-OCT-1997; 97US-0060837P.

XX 02-OCT-1997; 97US-0060838P.

XX 02-OCT-1997; 97US-0060839P.

XX 02-OCT-1997; 97US-0060843P.

XX 02-OCT-1997; 97US-0060866P.

XX 02-OCT-1997; 97US-0060874P.

XX 02-OCT-1997; 97US-0060880P.

XX 02-OCT-1997; 97US-0060884P.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Duan DR, Florence KA, Rosen CA, Ruben SM, Greene JM, Young P;

PI Ferrie AM, Yu G, Janat F, Ni J, Carter KC, Endress GA, Feng P;

PI Lafleur DW, Shi Y;

XX WPI; 1999-264022/22.

XX New isolated human genes and the secreted polypeptides they encode.

XX Disclosure; Page 341-342; 368pp; English.

XX This invention describes novel isolated human genes and the secreted proteins they encode. The products of the invention are useful for preventing, treating or ameliorating medical conditions, e.g. by protein or gene therapy. Also pathological conditions can be diagnosed by determining the amount of the new polypeptides in a sample or by determining the presence of mutations in the new polynucleotides. Specific uses are described for each of the 101 polynucleotides, based on which tissues they are most highly expressed in, and include developing products for the diagnosis or treatment of cancer, tumours, neurodegenerative disorders, developmental abnormalities and fetal deficiencies, blood disorders, leukemias, diseases of the immune system, autoimmune diseases, hepatic and renal disease, lymphomas, inflammation, allergies, Alzheimer's and cognitive disorders, schizophrenia, prostate disease, skeletal or cardiac muscle disorders, pulmonary disorders, transplant rejection, disorders involving osteoclasts such as osteoporosis, arthritis or malignancies, digestive/endocrine disorders, infections and AIDS. The human secreted proteins of the invention are represented in AAY07852-Y07993 and the encoding nucleic acids are represented in AAX37451-X37552

XX Sequence 149 AA;

SQ Query Match 7.7%; Score 662; DB 2; Length 149;

Best Local Similarity 96.1%; Pred. No. 9e-38;

Matches 122; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1250 HLSTLDRSVIWSKSLNARCKIKKGGDAENNVLCGDCGRGHHTYCVRPKLTVPDGVWF 1309

DB 12 HLSTLDRSVIWSKSLNARCKIKKGGDAENNVLCGDCGRGHHTYCVRPKLTVPDGVWF 71

QY 1310 CPECPKQRRRLSPQPSLESDESDVDSMGDEDEVDGDEEQSEEEYEVEQDEDD 1369

DB 72 CPECPKQRRRLSPQPSLESDESDVDSMGDEDEVDGDEEQSEEEYEVEQDEDD 131

QY 1370 SQEEVEV 1376

DB 132 SXEEXEV 138

RESULT 22

ABR41364

ID ABR41364 standard; protein; 1052 AA.

XX ABR41364;

AC ABR41364;

XX 02-JUN-2003 (first entry)

DT Human DITHP transcription factor.

XX Human; dithp; diagnostic and therapeutic polynucleotide; diagnosis;

XX cancer; cell proliferative disorder; autoimmune disorder;

XX inflammatory disorder; infection; hormonal disorder; metabolic disorder;

XX neurological disorder; gastrointestinal disorder; transport disorder;

XX connective tissue disorder; drug screening; proteome analysis;

XX gene therapy; antisense therapy; genotyping; transgenic animal; knock in;

XX disease model; toxicological testing; transcript imaging;

XX transcription factor.

OS Homo sapiens.

XX WO200297031-A2.

XX 05-DEC-2002.

XX 27-MAR-2002; 2002WO-US010056.

XX 28-MAR-2001; 2001US-0279619P.

XX 29-MAR-2001; 2001US-0280067P.

XX 29-MAR-2001; 2001US-0280068P.

XX 16-MAY-2001; 2001US-0291280P.

XX 17-MAY-2001; 2001US-0291829P.

XX 17-MAY-2001; 2001US-0291849P.

XX 19-JUN-2001; 2001US-0299428P.

XX 20-JUN-2001; 2001US-0299776P.

XX 20-JUN-2001; 2001US-0300001P.

XX (INCY-) INCYTE GENOMICS INC.

XX Daffo A, Jones AL, Tran AB, Dahl CR, Gietzen D, Chinn J;

PI Dufour GE, Hillman JL, Yu JY, Tuason O, Yap PE, Amshney SR;

PI Doughtery SC, Dam TC, Liu TF, Nguyen DA, Kleefeld Y, Gerstin EH;

PI Peralta CH, David MH, Lewis SA, Chen AJ, Panzer SR, Harris B;

PI Flores V, Marwaha R, Lo A, Lan RV, Urashka ME;

XX WPI; 2003-129518/12.

XX N-PSDB; ACC46304.

XX Novel human diagnostic and therapeutic polypeptide useful for identifying test compound which specifically binds to a polypeptide encoded by human diagnostic and therapeutic polynucleotide, and to induce antibodies.

XX Claim 27; SEQ ID NO 899; 591pp; English.

XX The invention relates to novel human diagnostic and therapeutic polynucleotides designated dithp (ACC46080-ACC46749) and to their encoded proteins (DITHP; ABR41136-ABR41812). The invention also relates to polynucleotide sequences at least 90% identical to the dithp cDNA sequences of the invention; recombinant vectors, host cells and transgenic organisms comprising a dithp nucleic acid sequence; the recombinant production of DITHP proteins; antibodies specific for DITHP proteins; microarrays comprising dithp nucleic acid sequences; methods of detecting dithp nucleotide and protein sequences; methods of screening for compounds which specifically bind a DITHP protein; and methods of assessing the toxicity of test compounds using a dithp hybridisation probe. Dithp nucleic acid sequences and DITHP proteins may be used in the diagnosis of a wide variety of conditions including cancer and other cell proliferative disorders; autoimmune or inflammatory disorders; bacterial, viral, fungal or parasitic infections; hormonal disorders; metabolic disorders; neurological disorders; gastrointestinal disorders; transport disorders; and connective tissue disorders. They may also be used to screen for modulators of protein activity or gene expression. DITHP proteins can additionally be used in analysis of the proteome of a tissue

CC clones are identical to any known human mRNAs. The present invention
CC describes an immunoassay to identify agonists and antagonists, as well as
CC antibodies, antisense molecules and siRNAs that can all be used to bind
CC to and modulate expression of the cDNA molecules. As such, these
CC molecules are useful for diagnostic markers or therapeutic targets for
CC the various diseases or morbid states. In particular, they are useful in
CC gene therapy for treating osteoporosis, neurological disease, Alzheimer's
CC disease, Parkinson's disease, dementia, short memory and various cancers,
CC as well as for maintaining equilibrium of sense or motor function, and
CC for treating emotional reaction, fear response and panic. Accordingly,
CC they exhibit osteopathic, neuroprotective, nootropic, antiparkinsonian,
CC cytoskeletal and tranquiliser activities. This polypeptide is a protein
CC encoded by a full length human cDNA sequence of the invention. NOTE: This
CC sequence is not given in the sequence listing of the specification but
CC can be obtained on CD-ROM from the European Patent Office, Vienna Sub-
CC office.

XX Sequence 524 AA;

Query Match 5.2%; Score 447.5; DB 8; Length 524;
Best Local Similarity 22.3%; Pred. No. 4.7e-22;
Matches 157; Conservative 95; Mismatches 196; Indels 255; Gaps 20;
QY 976 KTGP-LMSESTINIQPRDHSVQLPKPVHKPNR----WCFYSSCEQLDOLIBALNSRG 1030
DB 28 RAGDPGEMQPTGLGQ-----PKRRGRPPKMGSGWWIRDPMDLAKALHPRG 78

QY 1031 HRESALKETLLOEKSRICQAQLARSEKHFSDPKQDSKFTYSGRSSNAYDSQMCAB 1090
DB 79 IREKALHKLHNRDFLOEVLRLPSADPIF-----EPQLPAFOEGIMS--WSPKXTYE 131

QY 1091 KOLELRDLFDLTDRIYQGTGLGAIKYTDRIHRSALSGRYELLSEENKENGIIKTVN 1150
DB 132 --TDLAVLQWVEELEQRVIM-----SDLIQKGTWCPSPD-----STR 166

QY 1151 EDVEME--IDEQTKVIVKRLGIGIKTPTSTVSTNASTPOSVSVHYLAMALFOIEQG 1208
DB 167 EDLAYCEHLSQSDQEDITWRGR--GREGLAPQRKTNP-----LDLAVMLAALQN 215

QY 1209 IERRPLKAPLQASDGRGYKTVLDR-----WRESLSS 1241
DB 216 VERRYLRPL-----WPTHEVLEKALLSTPNGAPEGTTTBSYEITPRIRVWRQTLCRC 270

QY 1242 ASLSQVFLHSLDORSVIMWSILNARCKICRKGDAENMVLCDGCRGHHTYCVRPKLK 1301
DB 271 RSAAQVCLCLGLEKSTAWKSVKVTCLVCKRGDNEFLLLDCDCRGCHYCHRPKWE 330

QY 1302 TVPEGDFPCPCRPQRRLSFRQPSLESDVEDSMGDEDDVDGDEEGQSEEBEY 1361
DB 331 AVPEGDFCTVCLAQ-----VEGEFTQKPGFPKRGQKRSKY 368

QY 1362 EYVEDDSDQSEEEVSLPKRGQVRLPVKTRGKLSFSFSRGGQOQEGRPVPSRQSGTP 1421
DB 369 SLNFSQEG-----GR--RRVLLRGR-----ESPAAGPRYSSEG-- 400

QY 1422 KTTVSSKGRSLRKINSAPPTETKSLRTASRSHGCPLOADVVELLSRRKRGKRS 1481
DB 401 -----LSPSKRR----- 408

QY 1482 ANNTPENSPNPNFVRIATKSEQSRSVNIASKLSQSESKRCKRKRQSPPEPVTLGR 1541
DB 409 -----LSMRNHS----- 416

QY 1542 RSSRGGGVHLSAFLVLRHDSWPLKLVSKITQVPDYDIKKPIALNIIRKV 1601
DB 417 -----DLTFCEIILMEMESHAAPLEPNPLVSGYRIIKNPMDFSTMRRL 466

QY 1602 NKCEYKLASEFIDDIELMFSNCFEYNPNTSEAKAGTLOAFF 1644

DB 467 LRGVYTSSEFAADALLVFDNCQTFNEDDSEVGKAGHIMRRFF 509

RESULT 24

AAB93073
ID AAB93073 standard; protein; 708 AA.

XX AC AAB93073;

XX DT 26-JUN-2001 (first entry)

XX DE Human protein sequence SEQ ID NO:11898.

XX KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.

XX OS Homo sapiens.

XX PN EP1074617-A2.

XX PD 07-FEB-2001.

XX PF 28-JUL-2000; 2000EP-00116126.

XX PR 29-JUL-1999; 99JP-00248036.

XX PR 27-AUG-1999; 99JP-00300253.

XX PR 11-JAN-2000; 2000JP-00118776.

XX PR 02-MAY-2000; 2000JP-00183767.

XX PR 09-JUN-2000; 2000JP-00241899.

XX PA (HELI-) HELIX RES INST.

XX PI Ota T, Isoqai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

XX PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX DR WPI; 2001-318749/34.

XX PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.

XX PS Claim 8; SEQ ID NO 11898; 2537pp + Sequence Listing; English.

XX CC The present invention describes primer sets for synthesising 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesising polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH1833 to AAH18742 represent human cDNA sequences; AA892446 to AA895893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention

XX CC Sequence 708 AA;

XX SQ Query Match 4.4%; Score 381; DB 4; Length 708;

XX Best Local Similarity 25.2%; Pred. No. 3.2e-17;

XX Matches 161; Conservative 89; Mismatches 192; Indels 198; Gaps 24;

QY 420 SPANRRGRPP-----KRIHISQEDNVA--- 442

DB 157 SRMRGRGRPPNVGNVGNFLDNADAKLRKLKLAQQAQIKLRLKQKQEQARVAKEA 216

QY 443 --NKQTLASVRSKATKDKRDKLLKQEE----MKSLAFE-----KAKLREKADA- 484

Db 679 IQQLQKSIWAEKSIKMYVCQICRKGDNBEILLCLDGCDCGCHTYCHRPKITTIPOGDWFC 738

Qy 1311 PECRPKQ-----RCRLSFRQRPSPLESDEDEVDGSGGDEDEGEQSEEEYEVEQ 1365

Db 739 PACIAKASGQTLKIKLHVKGKKTNESK-----GKKVTLTGDTDEDSASTSSSLKR 791

Qy 1366 DEDD 1369

Db 792 GNKD 795

RESULT 26

ADL30971

ID ADL30971 standard; protein; 796 AA.

XX AC ADL30971;

XX XX

XX 20-MAY-2004 (first entry)

XX XX

DE Human protein encoded by a full length cDNA clone SeqID 3004.

XX human; medicine; signal transduction; glycoprotein; transcription;

KW oligo-capping method.

XX OS Homo sapiens.

XX EP1396543-A2.

XX FN 10-MAR-2004.

XX PD

XX PF 07-JUL-2000; 2003EP-00025638.

XX PR 08-JUL-1999; 99JP-00194486.

PR 11-JAN-2000; 2000JP-00118774.

PR 02-MAY-2000; 2000JP-00183865.

PR 07-JUL-2000; 2000EP-00114089.

XX (REAS-) RES ASSOC BIOTECHNOLOGY.

XX Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;

PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;

XX WPI; 2004-204755/20.

DR N-PSDB; ADL30970.

XX

PT New oligonucleotide primers (830 cDNAs) useful for synthesizing full

PT length human cDNAs.

XX Example 1; SEQ ID NO 3004; 1340pp; English.

XX

CC This invention relates to a novel primers useful for synthesizing full

CC length cDNA molecules that encode human proteins. Specifically, it refers

CC to secretory or membrane proteins that are potential therapeutic agents/

CC target molecules in the field of medicine, and in particular genes

CC encoding proteins that are associated with signal transduction.

CC glycoproteins and transcription. The present invention describes a method

CC for efficiently cloning a full length human cDNA from both the 5' and 3'

CC ends using the oligo-capping method. This polypeptide sequence is a full

CC length human protein of the invention.

XX

SQ Sequence 796 AA;

Query Match 4.3%; Score 368.5; DB 8; Length 796;

Best Local Similarity 20.6%; Pred. NO. 2.8e-16;

Matches 174; Conservative 104; Mismatches 247; Indels 319; Gaps 30;

Qy 793 LKEQEQKWKKEQK-----KEDEQNSTADISIGEER-----REDFDTSIE 834

Db 4 LRRDKVWVEGKLRRIIHAUKRTGSGGIDGSEQHPGLGTPPGRKRRKGDSYD 63

Qy 835 SKDTEQKELQDMFTEDDDPFGSHKRRGRKRGQNGFKETTRQEQINCVTRELLTADSEE 894

Db 64 DDDDDSDQGDDEDEED---KEDKKGKTD-----ICEDEDE 100

Qy 895 ---ALKQEHQKKEKLEKIQS-----AIACTNIFPLGRDRMYRRYWIFFPGLFI 943

Db 101 GDQAASVEELEKQLEKLSKQOSQYRRKLFDAHSLRSVMFGQDYYRRYWLPOCGGIFV 160

Qy 944 E--EDYSGLTEDMLLPFPSSFQNNVQSDPQVSTKTGPEPLMSESTSNIDQ----- 992

Db 161 EGMESGEGLEIAKEREKKAESVQIKEMFET--SGDSLNCSTNDHCEQKEDLKERDNT 219

Qy 993 -----PRDHSVQIPLKP----- 1003

Db 220 NLFQKPGSFKLSKLEVAKMPPESEVMTPKPNAGANGCTLSYQNSGKHSLSGVQSTAT 279

Qy 1004 ---VHK-----PNRWCFYSSCEOLDQLEALNSRGHRESAL-----KETLL 1041

Db 280 QSNVEKADSNLFWTSSGGPKG--FYSPLEN-DQLLKLTLTEKNQWFSLLPRTPCDDTSL 336

Qy 1042 QEKSRICAQLARFSEKFFHFDKQP----- 1067

Db 337 THADMSTASLVTPOSQPPSKSPPTPAPLGSSAQNPVGLNPPFALSPLQVKGVSMMGLQF 396

Qy 1068 -----DSKPTYSRG-----RSSNAYDPSQCAEQ----- 1092

Db 397 CGWPTGVVTGNIPFTSSVPSGLSGLSEGNSEFLTSNVASSKSESPPVQNEKATSAQP 456

Qy 1093 --LEL-RLRPF-----LLDIEDRIYQGTLGAI-----KV 1118

Db 457 AAEVAKPVPFPSPKPIPEEMQFGWRIIDP-----LKALKVLHLRIRKALKQK 510

Qy 1119 TDRH---IWRSALESRGYELLS--EENKENGIIKTV--NEDVEEMEIDEQTKVIIVKRLLG 1172

Db 511 IQHLDYITQACLKNDVAIIELNEENQVTRDIVENWSVEEQAMWDLVLQQVEDLE 570

Qy 1173 IKTEPPTVSTNASTPOSVS-----SVVHYL----- 1198

Db 571 RRVASASLQVKGWMCPEPASEREDLVYFEHKSFTKLCKEHGDEFTGEDESSAHLERKSD 630

Qy 1199 -----AMALFOIEQGIERRFLK--APLDASDSGRSYKTVLDRWRESILSSASISQVPLH 1250

Db 631 NPLDIATRLADLERNIERRIEDIAF-----GLRVWRRALRSARSAQVALC 678

Qy 1251 LSTLDRSVIWSKSIINARCKICRKGDAENMVLCDGCDRGHHTYCVRPKLTVPPEGWFC 1310

Db 679 IQQLQKSIWAEKSIKMYVCQICRKGDNBEILLCLDGCDCGCHTYCHRPKITTIPOGDWFC 738

Qy 1311 PECRPKQ-----RCRLSFRQRPSPLESDEDEVDGSGGDEDEGEQSEEEYEVEQ 1365

Db 739 PACIAKASGQTLKIKLHVKGKKTNESK-----GKKVTLTGDTDEDSASTSSSLKR 791

Qy 1366 DEDD 1369

Db 792 GNKD 795

RESULT 27

ABG97491

ID ABG97491 standard; protein; 1398 AA.

XX AC ABG97491;

XX XX

XX 16-DEC-2002 (first entry)

XX XX

DE Human NOVX12 protein.

XX

KW Human; NOVX; human disease; NOVX-associated disorder; cancer; addiction;

KW Hodgkin disease; Von Hippel-Lindau syndrome; Alzheimer's disease; stroke;

KW tuberculous sclerosis; hypercalcaemia; Parkinson's disease; depression;

KW Huntington's disease; cerebellar palsy; epilepsy; Lesch-Nyhan syndrome;

KW multiple sclerosis; ataxia-telangiectasia; leukodystrophy; anxiety; pain;

KW obesity; Crohn's disease; osteoporosis; inflammatory bowel disease;

KW infertility; inflammatory bowel disease; atherosclerosis; hypertension;

KW scleroderma; haemophilia; diabetes; pancreatitis; autoimmune disease;

KW asthma; arthritis; immunodeficiency; HIV; viral infection; neurogenesis;

KW bacterial infection; parasitic infection; graft-versus-host disease;
 KW cell differentiation; cell proliferation; haematopoiesis; wound healing;
 KW angiogenesis; single nucleotide polymorphism; SNP.
 XX Homo sapiens.

Key Location/Qualifiers
 FT misc_difference 139 /note= "Asp substituted by Gly as the result of a single
 FT nucleotide polymorphism (SNP)"

XX WO200272770-A2.

XX 19-SEP-2002.

XX 08-MAR-2002; 2002WO-US007283.

XX 08-MAR-2001; 2001US-0274281P.

XX 09-MAR-2001; 2001US-0274849P.

XX 12-MAR-2001; 2001US-0275235P.

XX 13-MAR-2001; 2001US-0275579P.

XX 13-MAR-2001; 2001US-0275601P.

XX 14-MAR-2001; 2001US-0276000P.

XX 20-MAR-2001; 2001US-0277239P.

XX 20-MAR-2001; 2001US-0277327P.

XX 20-MAR-2001; 2001US-0277338P.

XX 21-MAR-2001; 2001US-0277791P.

XX 22-MAR-2001; 2001US-0277833P.

XX 23-MAR-2001; 2001US-0278152P.

XX 26-MAR-2001; 2001US-0278894P.

XX 27-MAR-2001; 2001US-0279036P.

XX 28-MAR-2001; 2001US-0279344P.

XX 30-MAR-2001; 2001US-0280233P.

XX 02-APR-2001; 2001US-0280802P.

XX 02-MAY-2001; 2001US-0288148P.

XX 31-MAY-2001; 2001US-0294822P.

XX 31-OCT-2001; 2001US-0335302P.

XX 04-DEC-2001; 2001US-0338375P.

XX 07-MAR-2002; 2002US-00094465.

XX (CURA-) CURAGEN CORP.

XX Spytek KA, Vernet CA, Tcherven VT, Malyankar UM, Gerlach VL;

PI Li L, Zethusen BD, Patturajan M, Gusev VV, Kekuda R, Pena CEA;

PI Zhong M, Gangolli EA, Taupier RJ;

XX WPI: 2002-713508/77.

XX N-PSDB; ABS78735.

PT New NOVX polypeptides and polynucleotides, useful for preventing,
 PT diagnosing or treating NOVX-associated disorders, e.g. diabetes, multiple
 PT sclerosis, atherosclerosis, cancer, infections, osteoporosis or
 PT Parkinson's disease.
 XX Claim 1; Page 126-127; 266pp; English.

CC differentiation, cell proliferation, haematopoiesis, wound healing and
 CC angiogenesis, in gene therapy, in generation of antibodies that bind
 CC immunospecifically to NOVX substances for use in therapeutic or
 CC diagnostic methods. The nucleic acids are further used as hybridisation
 CC probes, in chromosome mapping, tissue typing, preventive medicine, and
 CC pharmacogenomics. The present amino acid sequence represents a human NOVX
 CC protein of the invention

XX Sequence 1398 AA;

Query Match 4.2%; Score 361.5; DB 5; Length 1398;
 Best Local Similarity 18.8%; Pred No. 2e-15;
 Matches 265; Conservative 184; Mismatches 457; Indels 505; Gaps 49;

QY 436 SQEDNVANKQTLASVRSKATKRDKLLKQEEKMSLAFAKAKLRK-ADAL 485

DB 175 SQDNNSSRES--PSLEDEETKKEETPKQEEQK-----ESEKMKSEQPMDLNRSTANVL 228

QY 486 E--AKKKEDEK------KRELKXIVEERLKKKEERLKV-523

DB 229 EETTVKKEDEKELVPLVIVKLEKPLPENEEKIIKEESDPFKNVPIKVEKCR 288

QY 524 -----REKER-----EKLREKKYVEYLQWQSKP 548

DB 289 DPKTKSMKPVQAQEPERIEFGNKKSSHEITEKTEKLNQDQAKIPLKREIKL 348

QY 549 REDME-----CDDLKELPEPTPVKTRLPPIFGDALMVLEFLNAFGLDQEFDPG 601

DB 349 SDDFDPVYKPLCKSV-----TPTKEFLKDEIKQEE-----ETCKRISTITALGHEGKQL 398

QY 602 VTLEVLREALVGNSEGPLCELLPFFLTAIQAIAEEEEVAKEQLTADTVGCSLSLD 661

DB 399 VNGEVSDEKAPNFKEPI-ETKFF-----ETKEESYSPSKDRNIITEGNGTESLN 448

QY 662 --LDSCITLSEILR-LHILASGADVTSANAKYQKGGFDATDDACMELRLNPSLVKL 718

DB 449 SVITSMKTGELEKETAPLRKQAD--SSISVLEIHSOKAQIEPDPEMETSLDSEMAKDL 507

QY 719 SSTSVYDLTPGCKMKILHALCGKLLTLVSTRDF---IEDYVDILRQAQKEFRELKAEQRH 775

DB 508 SSKTALSSTESTMK-----GEEKSPKTKDKRPPILECLEKLEKSKTKFLDKDAQRLS 561

QY 776 KEREAAARIRKKEEKLKEQEQMKKEQ-----EKL---KEDQORNSTADISIGE 823

DB 562 PIPEE---VPKSTLESEKPGSPEAAETSPPSNIIDHCEKLAKEVEVCQSTVGGOS 617

QY 824 EEREDFTSIESKDEQKELD--QDMFTDEDDPSHKGRRGKGGQNGKFEFTQEQIN 881

DB 618 VKKVDLETLEDSEFTKVMNDNLDNAQTSGLIEPSETK---GSMQKSKFK-----YK 666

QY 882 CVTRELTLADEEALKQEHQKKELEKIOSAIACTNIFPLGRDRMYRRYWFPSIPGL 941

DB 667 LVPSEETASENTETTSERQEGIKLTIRIS-----RKKK-----PDSPPK 708

QY 942 FIEEDYSGLTEDMLLPRESSFNQNVQSDPQVSTKTSPLMSESTSNIDQGRDRHSVQLP 1001

DB 709 VLE-----PENQKETEKEEK--TNVGRTLR-----RSPRIS 739

QY 1002 KPVHKPNWCYSCQEQDLQIEALNSRGHRESALKETLLQKESRICQALRPFSEKHFH 1061

DB 740 RPTAK-----VAEIRDQKADKKRGEGEVEEESTALQK-----TDKKEI 779

QY 1062 SDKPOPDSPKPTVSRGRSSNAYDPSQMAEKQLELRDLFLDIEDRIYQGTGLGAIKVTRD 1121

DB 780 LKKSEKDTNSKVK-----VVKPKGK 799

QY 1122 HIWRSALSGRYEYLLS--EENKENGIIKTWNEDVEEMIDEQTKVIVKDRLLGIKTETPST 1180

DB 800 VRWTCSTRGRWYKYSNDESGSGSEKSSAAASEEEKESEEAIALDD-----847

QY 1181 VSTNASTPQSVSSVVHYLAMALFQIEQGIERRFLKAPLDASDSGRSKYKTVLDWRRESLLS 1240

DB 848 -----847

QY 1241 SASLSQVFLHLSTLDRSVIWSKSIILNARCKICRKKGAENMVLCDGCDRGHHTVCVRPKL 1300
 DB 848 -----DEPKCGGLPNHPELILLCDSGSHHTACLRPL 882
 QY 1301 KTVPEGDFWPCPCRPQRCRL-----SFRQPSLESD-----VEDSMGE 1343
 DB 883 MIIPDGEWFCPCQHKLLCEKLEQLDLVALKKERAERKRLVVGVSISINIIPPQ 942
 QY 1344 DDEVGDDEEGQSEEEVEV-----EDEDSDQ-----EDEDSDQ----- 1372
 DB 943 EPDFSEDEEKKQSKSKANLERRSTRTRKCI SYRDFEFAIDAEIDDIKEADGG 1002
 QY 1373 -----EEVSLPKRGPQVRLPVKTRGKLSSPS----- 1401
 DB 1003 VGRGKDITITGRGKIDSTILDERKENKPKQRAAAARRKRLNLDLSDNLDEES 1062
 QY 1402 -----SRGQQE---PGRYPSRQOQSTPKTTVSK--TGRSLRKINSAPPTETKSLRIA 1450
 DB 1063 EDEFKISDGSQDEFVSDENPDESEEDPNDSDTDTCFCSRLRHPSRPMQSRRLRK 1122
 QY 1451 SRSTRHSHGPIQ-----ADVVELLSPPRRKRGKRSKANSNTPENSFN 1492
 DB 1123 TPKKKYSDDDEESEEENSRDSEDFSDDFVE-TRRRSRNRQKQINYNKEDSED 1181
 QY 1493 PNFRVIATKSEQSRSVNIASK--LSLOES-----ESKRRCRKR----- 1529
 DB 1182 G-----SOKSLRRGKEIRVRHKLSSSEESYLSKNSDEDELAKESKRVKRGST 1236
 QY 1530 ---OSPEPSVTLGRSSRGQGVHLSAFQLVVELVRHDDSWPFLKLKLSKIQVPDYD 1586
 DB 1237 EYSEADEEEBEGKPSRKR---LHRIETDEESCDNAHGDAQPDSQPRV-LPSEQE 1292
 QY 1587 IIKPIALNIIRKVNCKEYKLAGEFIDIE 1617
 DB 1293 STKXP-----YRIESDEEDFE 1309
 RESULT 28
 ID AAB83348
 XX AAB83348 standard; protein; 1400 AA.
 AC AAB83348;
 DT 09-OCT-2001 (first entry)
 DE AAP-2 protein sequence.
 KW Apoptin-associated protein; p53-independent apoptosis induction;
 KW cell proliferation; cell death; cancer; autoimmune disease; Vector.
 XX Unidentified.
 OS
 FH Key Location/Qualifiers
 FT Misc-difference 1184..1201
 FT /note= "encoded by GGA"
 XX
 FN EP1118662-A2.
 PD 25-JUL-2001.
 PF 17-JAN-2001; 2001EP-00200163.
 PR 17-JAN-2000; 2000EP-00200169.
 PR 07-APR-2000; 2000EP-00250118.
 PA (LEAD-) LEADD BV.
 PI Noteborn MMH, Danen-Van Oorschot AM, Rohn JL, Weiss B;
 DR WPI; 2001-477075/52.
 DR N-PSDB; AAF87096.
 XX

PT Nucleic acids encoding Apoptin-associated proteins, useful for the
 XX prevention, diagnosis and treatment of cancers and autoimmune disorders.
 PS Claim 12; Fig 8; 32pp; English.
 XX

This sequence represents the AAP-2 protein sequence. The invention
 relates to nucleic acids encoding Apoptin-associated proteins. The
 CC nucleic acid, vector containing it, host cell and/or the encoded protein
 CC may be used to induce p53-independent apoptosis. This may further
 CC comprise using a nucleic acid encoding Apoptin or functional (or
 CC functional equivalent/fragment). They may be used in this way for the
 CC treatment of a disease in which enhanced cell proliferation or decreased
 CC cell death is observed, especially a cancer or autoimmune disease. The
 CC nucleic acids may also be used for detecting the presence of cancer cells
 CC (or cancer-prone cells) in a sample, by transfecting cells in the sample
 CC with the nucleic acid or vector, and determining the percentage of
 CC apoptotic in the sample. This may further comprise determining the
 CC intracellular localisation of a proteinaceous substance derived from the
 CC nucleic acid or vector in the cells. The presence of the proteinaceous
 CC substance is detected by immuno-staining the cells with the antibody.
 CC They may also be used to identify a putative cancer-inducing agent (e.g.
 CC a genome or functional fragment) by contacting a cell sample with a
 CC candidate agent and detecting the presence of cancer cells (or cancer-
 CC prone cells). They may also be used to identify a putative effector of the
 CC activity of protein, by binding it with a candidate effector and
 CC determining it's binding

XX Sequence 1400 AA;

Query Match 4.2%; Score 361.5; DB 4; Length 1400;
 Best Local Similarity 18.8%; Pred. No. 2e-15;
 Matches 265; Conservative 184; Mismatches 457; Indels 505; Gaps 49;
 QY 436 SQEDNVANKOTLASYRSKATKRDKLLKQEBMKSLAFKAKLKREK-----ADAL 485
 DB 177 SQDSSRES--PSLEDEETKKEETPKQEBQK---ESKMKSEOPMDLENSTANVL 230
 QY 486 E--AKKKEKEDKEK-----KREELKIVVEERLKKKEERLKV----- 523
 DB 231 BETTVKKEDEKELVKLPVIVKLEKLPENEEKKIKEESDSPKENVKPIKVEVKECA 290
 QY 524 -----REKER-----EKLREERKRYVEYLKQMSKP 548
 DB 291 DPKDTKSSMEKPVQAEPEPIEFPGCNKSSHEITEKSTEETEKLNQDQAKIPLKKREIKL 350
 QY 549 REDME-----CDDLKELPEPTPVKTRLPPEIFGDALMVLEFLNAGELFDLODEFPDG 601
 DB 351 SDDPDSVPKGLCKSV-----TPTKFLKDEIKQEE---ETCKRISTITALGHEGKQL 400
 QY 602 VTLEVLVEALVGNDSGLPCLELFFFLTAIFAABEEVEVAKQLTDADTKGCSLSKLD 661
 DB 401 VNGEVSDEVPAPNFKTEPI-ETKFI-----ETKEESYSPSKDRNIITEGNGTESLN 450
 QY 662 --LDSCTLSEILR-LHILASGADVTSANAKRYOKRGGFADTDACMELRLSNFSLVKKL 718
 DB 451 SVITSMKTGELEKETAPLRKQAD--SSISVLHSHQAKIEEPDPPEMETSLLDSSEMAKDL 509
 QY 719 SSTSVYDLTPGERKMKILHAGCLLTLVSTRDF---IEDYVDILRQAKQEFRELKAEQHR 775
 DB 510 SSKTALSSTESCTMK-----GEEKSPKTKDKRPPILECLEKLEKSKTKTFLDKDAQRLS 563
 QY 776 KEREEAARIRKKEEKLKQEQKKEKQ-----EKL---KEDEQRNSTADISIGE 823
 DB 564 PIPEE-----VPKSTLESEKPGSPAEATSPSPNIIDHCEKLASEKVEECOSTSVVGQS 619
 QY 824 EEREDFTSIESKDTQOKELD--ODMFTEDDDDPGSHKRGRRGRKGQNGKFEFTRQEQIN 881
 DB 620 VKKYDLETLKEDSEFTKVMNDLNAQTSGIEEFSETK---GSQKSKFK-----YK 668
 QY 882 CVTRELITADEEALKQEHQKKEKLEKLEKTSQAIACNIIFFLGRDMRVRRTWIIPSIPL 941
 DB 669 LVPEETTASENTEITSERQKEGIKLATIRISS-----RKKK-----PDSPPK 710

CC polypeptides; expression vectors and host cells comprising a TAT nucleic
 CC acid; an antibody specific for a TAT polypeptide; a peptide or organic
 CC molecule which binds to a TAT polypeptide; fusion proteins comprising a
 CC TAT polypeptide; and methods and compositions for the treatment or
 CC diagnosis of cancer in mammals. TAT polypeptides, nucleic acids,
 CC antibodies, antagonists, binding molecules and compositions are useful
 CC for diagnosing or treating a cell proliferative disorder associated with
 CC increased TAT expression, particularly cancers such as breast cancer,
 CC colorectal cancer, lung cancer, ovarian cancer, liver cancer, bladder
 CC cancer, pancreatic cancer, cervical cancer, cancers of the central
 CC nervous system, melanoma and leukemia. TAT nucleic acids may further be
 CC used as hybridisation probes, in chromosome and gene mapping, in
 CC chromosome identification and in gene therapy. The present sequence
 CC represents a TAT polypeptide of the invention
 XX
 SQ Sequence 1584 AA;

Query Match 4.0%; Score 343.5; DB 8; Length 1584;
 Best Local Similarity 20.5%; Pred. No. 4.4e-14;
 Matches 278; Conservative 210; Mismatches 514; Indels 355; Gaps 51;

QY 307 DGETIISDDSTQ-----SCSPQNGKKDAIDPLLFKYVQPTKKELHESAIK 358
 DB 104 DGKESLLQDRQEQRRFEPDRDQLEBEPGQR-----RQKQEQRELAE 151
 QY 359 ATQISRRKHLFSRDKL---FLKQCEPQ--GVIKIKASSLSTYKIAEQDFSYFPDP 414
 DB 152 EQSEKQERLEQDRQRDEELWRQEQWQEREERAEQLOSCKGHTE---EFPDDE 208
 QY 415 PTFIFSPAN-RRGRPPK-RHISQEDNVANKQTLASYSKAT-----KSRDKLLKQEE 466
 DB 209 QLRRELLELRKGRQEQRRQDRVFQEESEKWKRETVLKKEEKREWLKLE 268
 QY 467 MKSLAFKAK---LKREKADALEAKKKEKEKEK-----KREELKIV 506
 DB 269 EERREQERQQLRQEQERREQLKQEQEERLQQLRSEQLRQEQERREQLKQEQ 328
 QY 507 BEERLKKKEERLKVREKEREKUL--REE-----KKYVYLYKQMSKPREMECDL 557
 DB 329 BEKLEQERREQLKQEQEERDQLKREERROQLKQEQEERLEQLK-REEVERLEQ 387
 QY 558 KELPEPTVTKLPEIIFGDALMWLFNAGE---LFDLQDEFPDGVTLVLEAL--- 611
 DB 388 EERREQ--LKREEPEERQQLKSEEQEERQQLRQEQERREQLKQEQEERLEQ 446
 QY 612 VGNDSQGLCLLFFFLTAIFAQIAEEREEVAKQETDADTKGCSLKSLDLDSCSL 671
 DB 447 LKREHEERRE-----QELAEQEQARERIKSRIPN----- 479
 QY 672 RLHILASGADVTSANAKYQKRGGFDTDDACBELRLSNFLVKLSSTSVYDLTPGEK 731
 DB 480 -QWOLESEADAROSKVYSRPRKQEQRRREQ--BEKRRRRESELQWQEEERAHQQQEE 537
 QY 732 MKILHALCGKLLTLVSTDF-----IEDYDLRQ-----AKQEPFELKAEQHRK-- 776
 DB 538 QR-----RDFTWQWQAEKSEGRQRLSARPPPLRQERQRLAEERQORE 582
 QY 777 -----EREAAAIRKKE-----EKLKEQEQKMKQEKLEDE 811
 DB 593 QRFLPEEKEQRQRREKEREKELQFLEEQQLRERAAQQLQEEEDGLQEQRRRQOE 642
 QY 812 QRNSTADISIGEEEREDFTSIESKDTQKEL--DQDMFTEDDDPGSHKRGGRKGRQN 869
 DB 643 QRDDQKWQWLEERKRRRHTLYAKPALQEQQLRKEQQLQEEBELQREERKRRRQOE 702
 QY 870 GFKEFTREQINCVTRELTLDADEALKQEHQ---RKEKELLEKIQSAIACNTFPIG-- 924
 DB 703 --ROYREEQLEQEEQLREERKRRRQERQYRKDKKLQKEEQ-----LGE 752
 QY 925 -----RDRMYRYWIFPSIPGLFIEDYSLGTEDMLPRPSSFQNNVQSDQPOVSTK 976
 DB 753 PEKRRRQERKKYR-----EEELQEEEQQL-REERKRRRQEWERYKK 798

QY 977 TGEPLMSESTNIOGPRDHSVQVLPKPVHKPNRMCYSSCQOLDQLIILMNSRGHRESAL 1036
 DB 799 --DELQEEQLLREERKRLQ-----ERQYREEBELQEEBQL--LGERETR 846
 QY 1037 KETLQESRICAOLARSEKHFSDKPDSPDSTYSGRSSNAYDPSQMCACQLELR 1096
 DB 847 RQELERYRKEEBELQEEBQL-----REBPKRRRQERQCRNERQFQLEDSQLDR 902
 QY 1097 -----LRDLFDIEDRYQGTGAIKVTDRIH-----WRSALSGRYELLSEENKENG 1144
 DB 903 SQSQDLQHLQEQERDEQERRRQWRDRHFPEEQLEREBEQEKAKRDRKSQEQ-- 960
 QY 1145 IIKTVNEVMEIDEQT--KVIKDRLLGKTTPTSTVSTNASTPQSVSSVVHYLAMAL 1202
 DB 961 ---LLREEREKRRRQETDRKFRREEQLQREEQPLR----- 995
 QY 1203 FQIEQGIERRFLKAPLDASDSGRSVKTVLDRWRESLLSSASLSQVFLHLSTLDRSVWSK 1262
 DB 996 ---RQERDRKPREELRHQEQGRKFLBEEQLRQ-----ERERKFLK 1035
 QY 1263 SILNARCKIKRKGDAENMVLCDGCDRGHTTYCVRPKLKTVPEGDFCPCRPKQRCRL 1322
 DB 1036 EEQQLRCQ-----EREQQLAQD-----RDR 1055
 QY 1323 SFQRPSLEDEDVEDSMGDEDDVDGDEEGQSEEBEYVEQDEDDQSEEBEVLKRG 1382
 DB 1056 KFREEEQQLSRQDRKFRREEQVVRQERKERKLEEQQLRQERHRKFRREEQLLOBRE 1115
 QY 1383 RPOVRLPVKTRGKLSLSSPSSRGQEQEPGRYPSRQOSTPPTVSSKTGRSLR--KINSAP 1440
 DB 1116 EQQLHROERDRKFLBEEQQLRRQERD-----RKFREQLRQOE 1153
 QY 1441 PTETKSLRIASRSTRSHGHPQADVFVLLSPRRKRRGRKSANNTPENSPNPNRVAT 1500
 DB 1154 P-ERKFLBEEQQLRQER--QRKFLQEEQQLRQERQQRQDRD-----KFR----- 1199
 QY 1501 KSSQOSRVNTASKLSLQSESKRCR-----KROSPE 1533
 DB 1200 -EEQLQREBEQQLSRQERDRKFLBEEQKVRQOE 1235

RESULT 31
 AAB63263
 ID AAB63263 standard; protein; 244 AA.
 XX AC AAB63263;
 XX AC AAB63263;
 XX DT 26-MAR-2001 (first entry)
 XX Human breast cancer associated antigen protein sequence SEQ ID NO:625.
 DE Human; breast cancer; gastric cancer; prostate cancer; diagnosis;
 KW cancer associated antigen; cytostatic; cancer vaccine.
 XX Homo sapiens.
 OS WO200073801-A2.
 FN 07-DEC-2000.
 PD 26-MAY-2000; 2000WO-US014749.
 PF 28-MAY-1999; 99US-0136526P.
 PR 10-SEP-1999; 99US-0153454P.
 XX (LUDW-) LUDWIG INST CANCER RES.
 FA Obata Y;
 PI WPI; 2001-025274/03.
 DR Nucleic acids encoding breast, gastric and prostate cancer associated
 PT antigen precursors, useful for diagnosing and treating a condition

PT Characterized by expression of an abnormal amount of a protein, e.g. cancer.

XX

PS Example 1; Page 488-489; 799pp; English.

XX

CC AAF22422 to AAF22626, AAF22627 to AAF22773 and AAF22774 to AAF23014 represent nucleotide sequences encoding human breast, gastric and prostate cancer associated antigen precursors (CAAP) respectively.

CC AAB63232 to AAB63467, AAB63468 to AAB63721 and AAB63722 to AAB63970 represent human breast, gastric and prostate CAAP protein sequence respectively. CAAPs have cytostatic activity and can be used in the production of cancer vaccines. The human CAAP proteins, peptides, nucleic acids or anti-CAAP antibodies are useful for diagnosing and treating a condition characterised by expression of an abnormal amount of a protein, e.g. cancer

XX

Sequence 244 AA;

Query Match 3.9%; Score 138; DB 4; Length 244;

Best Local Similarity 24.0%; Pred. No. 6.2e-15;

Matches 92; Conservative 48; Mismatches 81; Indels 162; Gaps 8;

QY 1243 SLSQVFLHSLTDRSVMSKILNARCKIKRKGDAENMVLCDGDRGHHTYCVRPKLKT 1302

Db 4 SAAQVALCTIQLOKSIAMKSIKVKYQICRKGDNELLLLCDGCDKCHTYCHRPKIT 63

QY 1303 VPEGDFWCFCEPRKQ-----RCRLSFRQSLSDSDVDSMGDEDDVGDSEEGQSE 1357

Db 64 IPDGDWFCFACIAKASGQTLKIKLHVKGKTNESKK-----GKVTLTGDT----- 111

QY 1358 EEEVEVEDEDDSDQEEBVSILPKRGPQVRLPVKTRGKLSLSSFSRSGQGPGRYPERSQ 1417

Db 112 -----DED----- 114

QY 1418 QSTPKTVSSKTRSLRKINSAPPTETKSLRIASRSTRHSRGLQADVVELLSPRKRR 1477

Db 115 -----SASTSSSL----- 124

QY 1478 GRKANNTWPNPFPNFRVATKSSQSRVNIASKLSQESKRKRRCRQSPSPV 1537

Db 125 GNKD-----LKKRMEENTIN-----LSKQESFTSVKKPRDSDK----- 160

QY 1538 TLGRSSRGQGVHLSAFEOLVVELVRHDSWPLKLVSKIQVDDYDIKKPIALNII 1597

Db 161 -----DLALCSMLTETMETHEDAWFLPVLNKLVPGYKKVKKPWFSTI 206

QY 1598 REKNKCYKLASEFIDDIELMF 1620

Db 207 REKLSSGQYPNLETFALDVLVLF 229

RESULT 32

ID ABG05850

XX ABG05850 standard; protein; 2246 AA.

AC ABG05850;

XX

DT 13-FEB-2002 (first entry)

DE Novel human diagnostic protein #5841.

XX

KW Human; chromosome mapping; gene mapping; gene therapy; forensic; food supplement; medical imaging; diagnostic; genetic disorder.

XX

OS Homo sapiens.

XX

FN WO200175067-A2.

XX

PD 11-OCT-2001.

XX

PF 30-MAR-2001; 2001WO-05008631.

XX

PR 31-MAR-2000; 2000US-00540217.

PR 23-AUG-2000; 2000US-00649167.

XX (HYSE-) HYSEQ INC.

XX Drmanac RT, Liu C, Tang YT;

PI WPI; 2001-639362/73.

XX N-PSDB; AAS70037.

XX New isolated polynucleotide and encoded polypeptides, useful in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits and to assess biodiversity.

PT

PT Claim 20; SEQ ID NO 36209; 103pp; English.

XX

CC The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic amino acid sequences of the invention. Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX

Sequence 2246 AA;

Query Match 3.8%; Score 326.5; DB 4; Length 2246;

Best Local Similarity 19.6%; Pred. No. 1.2e-12;

Matches 335; Conservative 243; Mismatches 612; Indels 523; Gaps 75;

QY 21 PGSPLS-----LPVSTGCKSHRVANKVEARSEKLLPTALPSEPK-----VDQ 65

Db 341 PGTPESKATSCFPPTPRDRHEGRKQSTTEAQAAPASTKQADRRQSMFSLNTPK 400

QY 66 KLPRSSRRSGGGTQFPARSRAVAAGEAAAGAPERGSPGLGRRVSPR--CLCSGEGG 123

Db 401 KLGNLSLRGA-----SKALSKASPNTRS--GTRSPRIATTASAT 442

QY 124 QVAVGVIAKGRGRGRD-----GSRRAFGGEMPLLLHRKPFVRQKPPADLR 169

Db 443 AAAGATPRAKGLGQDDVTLDNLFRVHEHTGCVAGVQGRQLQH-----QGLAQLR 495

QY 170 PDEVFYCKVTNEIFRHYDDFFERTILCNLSLVNSCAVTGRPLTYQEALESEKARQNLQ 229

Db 496 -----TCAPTFPAG-----AEGKERYGFQ 514

QY 230 SFPEPLIIPVLYLTSLTHRSRLHEICDDIPAYVKDVFVEETVIRNNGARLOCTILEV 289

Db 515 E-----VHPFYVVVIA-ENRKHPSSPECLV-----SAQKVLK-----GSELELAKWTM 557

QY 290 LPPSHQNGFANGHVNSVDGETIIISDSDSDSCSFQNGKKKDAIDPLLFYKVQPTPK 349

Db 558 LLLVH-----STMS-----SKSPRDWEQPEYKIQ----- 582

QY 350 ELHESAIVKATQISRRKHLFSRDKLFLKQHCSPQEGVIKASSISTYKIASQDFSYF 409

Db 583 ---ELAVILKFLVDHEDGLNLEDLENFLQ-----KAPVPST-----CSST 620

QY 410 FPDDPPTTFISPANRRRGRPPKRIHISQEDNVANKQTLASVRSKATKRDKLLKQEMKS 469

```

Db 621 FEE-----LSPPSQAKREIRFLQKVASSSSGNFLS-GSPASPMGDIL-::
Qy 470 LAFERAKLREKADALEAKKKEKEDKREELKIVEE-----ERLKKKEK 517
Db 669 PQFQMRLLKQLAD---ERSNRDELELAENRKLTKDAQIAMQOQIDRLALLNEK 724
Qy 518 ERLKVEREKERKLEBEKKKYVEYLKQMSKPREDMEDCDLKPBPPTVKTRLPPEIFGD 577
Db 725 QASPLEKPELBEELDKNESLTMRLHETLK-----QCODLKTEKSMQDKINQLSEEND 779
Qy 578 -ALMVFLEFNAGFLDQDEFPD---GVTLEVLSE-----AL-----VGN 614
Db 780 LSKFEREFASHLQQLADALNEUTEHSHKATQEWLEKQALEKELSAALQDKARAKGDLGN 839
Qy 615 DSEGLCELLFFFLTAIFAQIAEBEEVAKQETDADTKGCSLKSGLDLDSC--TLSEILR 672
Db 840 KQMGPMFADVYGLGNLKL-EEKNEILQKLSQLEHLSQLQDNPPOKEGEVLGDVLQ 898
Qy 673 LHILASGADVTAN-----AKRYQKGGFDPATDDACHMELRLSNPS 713
Db 899 LETLQEAATLAANNQOLQARVEMLETERGQOEAKL-LAERGHFE-----EEKQLSS 950
Qy 714 LVKLSSTSVYDLTSG-----ERWKILH-----ALCGKLLTLVSTEDFIEDVILR 760
Db 951 LITDLQS-SISNLSQAKLEQSAQHGARLTAQVASLTSUETLTLNAT---IQOODQELA 1006
Qy 761 QAKQEFRELKAE--QHRKEREAAAIRKKEE---KLKEQKQMKKEQKLEKDEQORNS 815
Db 1007 GLKQAKKEQQAQTLQOOEQASXLRHQVEQLSSSLKQKEQQLKVAEK-QEATRODH 1065
Qy 816 TADISIGBEERDFDTSIESKTEQKELQDMFTEDDDPGSHKRGKRGQNGKFEFT 875
Db 1066 AQOLATAABERE---ASLRERDAALKOLEA-----LEKE 1096
Qy 876 ROEQINCVTRELLTADDEEALKQ---EHORKEKELLEKIQSAIACNTNIFPLGRDMYRR 931
Db 1097 KAAKLEILOQQLQVANEARDSAQTSVTQAREKASLRSKVESLOAC-----1142
Qy 932 YWIFPSIFGLFEEDYSGLTEDMLPSPSSFQNNVQSDQPVSTKTGPELMSSESTNIDQ 991
Db 1143 -----VETARQEQHEAQAVAELELQ-----LRSEQ-----1168
Qy 992 GPRDSHVLPKPVHPKPNWCFVSSCEQLDQLTEALNSRGRHSALKETLLOKSKHICAO 1051
Db 1169 -----QKATEKER-----VAQEKDQLQEQALQKESLKVTKGSLSEKRAADAL 1213
Qy 1052 ARFSEKHFSDKQPDQSKPTYSRGRSSNAYDPSQWCAEKQLELRLDF--LLDIEDRIY 1109
Db 1214 ---EEQQRCTSELKATRSLSVQHKRERKELEERAGRKGLARLQQLGEAQAQATEVL 1269
Qy 1110 QGTLCAIKVTDRIHWSALES-----GRYELLSEENKENG1-----IKTVNEDVEE 1155
Db 1270 RREL-AEAMAAQHTAESECEQLVKEVAWREYEDSQEEAQYGMFQEQMLTKECEK 1328
Qy 1156 --MEIDEQTKVIVKDRLLGIKTETPTSTVSTNASTPQSVSVVH-YLAMALFOIEQIGIRR 1212
Db 1329 ARQELQE-----AKEVAGIESHSELOJSRQ---QNELAELHANLALQVQVE---K 1375
Qy 1213 FLKAPLDASDGRSYKTVLDRWRESLLSASLSQVFLHLSTL-DRSVTWKSLNARCK- 1270
Db 1376 EVRAQKLADD-----LSTLQEKWAATSKEV--ARLET 1405
Qy 1271 ICRKKGDAENMVLDCDGRGHHTYCVRPKLKTVPEGDMWFCPCRCRKPORCRRLSFPORSL 1330
Db 1406 LVRKAGEQOET-----ASRELVKEPAPAGDRQPEWLEEQOQOQPCSTQ-AAL 1451
Qy 1331 ESDDEVDSMGDEDDVDGD--EEEQGSEEEYEVEQDEDDSQEEEVS--LPKRGRRQV 1386
Db 1452 QAMEREAEQMGNELERLAALMESOGQOQEEER-----QQQREVARUTQERGRAQA 1502
Qy 1387 RLPVK--TRGKLSFSFSRGGQOQEPGRYPFSRSQQSTPTTSSKTKGR-----SLRKINA 1439

```

```

Db 1503 DLALEKAAAEMLQNALNEQ-----RVFATLQALAHALTEKSGKQDELAKLRGLEAA 1559
Qy 1440 PPTETKSLR-----IASRSTRSHGPIQADVVELLSPRRKRGRKS--ANNTPEN 1488
Db 1560 QIKLEELURQTVKQLKEQLAKKEHAGS-----SGAQSEAGRTEPT 1602
Qy 1489 SPNPNFNVIAATKSEQ-----SRVNIASKLSLOESESKRRCRKRQSEPSVPV 1537
Db 1603 GPKLEALRAEVSKLEFQCCQKQOQADSLERSLE-AERASRAERDSALETLQOLEEKAQ- 1660
Qy 1538 TLGRSSRGQGVHLSAFQELVVELVRHDDSW 1570
Db 1661 ELGHSQSALASAQRELAARFKVQDHSKAEDW 1693

RESULT 33
AAMI4838
ID AAMI4838 standard; protein; 572 AA.
XX AAMI4838;
XX 12-OCT-2001 (first entry)
XX Peptide #1272 encoded by probe for measuring cervical gene expression.
XX Probe; human; microarray; gene expression; cervical epithelial cell;
XX cervical cancer.
XX Homo sapiens.
XX WO200157278-A2.
XX 09-AUG-2001.
XX 30-JAN-2001; 2001WO-US000670.
XX 04-FEB-2000; 2000US-0180312P.
XX 26-MAY-2000; 2000US-0207456P.
XX 30-JUN-2000; 2000US-00608408.
XX 03-AUG-2000; 2000US-00632366.
XX 21-SEP-2000; 2000US-0234687P.
XX 27-SEP-2000; 2000US-0236359P.
XX 04-OCT-2000; 2000GB-00024283.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-489901/53.
XX Human genome-derived single exon nucleic acid probes useful for analyzing
XX gene expression in human cervical epithelial cells.
XX Claim 27; SEQ ID NO 19664; 487pp; English.
XX The present invention relates to human single exon nucleic acid probes
XX (SENP; see AAI10068-AA128459). The present sequence is a peptide encoded
XX by one such probe. The SENPs are derived from human HeLa cells. The SENPs
XX can be used to produce a single exon microarray, which can be used for
XX measuring human gene expression in a sample derived from human cervical
XX epithelial cells. By measuring gene expression, the probes are therefore
XX useful in grading and/or staging of diseases of the cervix, notably
XX cervical cancer. Note: The sequence data for this patent did not form
XX part of the printed specification, but was obtained in electronic format
XX directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX Sequence 572 AA;
XX Query Match 3.7%; Score 318.5; DB 4; Length 572;
XX Best Local Similarity 24.8%; Pred No. 5.2e-13;
XX Matches 124; Conservative 89; Mismatches 173; Indels 115; Gaps 20;
Qy 429 PPKRIHI-----SQEDNVANKQTLASYRSKATKRDKLLKQSEMKSLAFKAKLRKAD 483

```

Db 175 PPAALHLIAYYKKNDRKDSALSCVSKTA---RLSSDDRARLPPEELSLVQKRYE 230
Qy 484 ALEAKK-KEDDEKKREELKVIIEERLKKKEKERLKVREKERELREKKRYEYL 542
Db 231 LLEHKRWASMSERKEYLKK--KREELKK-LKAKAKERREKEMLERLEKQKRY---- 283
Qy 543 KQMSKPRDMCEDDLKELPEPTPVKT--RLPPEIFGDALMVLFLNAGFELFDLQDEPPD 600
Db 284 -----EDQELTG-KNLPAPRLVDTPEGLNTLFGDVAMVVEFLSCYSGLLLPDAQYP- 334
Qy 601 GVTLEVLBEALVGNDSGELCELLFFLTAIFQAIABE--BEEVAKEQLTDADTKGCSLK 658
Db 335 -ITAVSLMEAL-SADKGG-----FLYNRLVILLQTLLODEIAE---DYGELGMKLS 382
Qy 659 SLDDSDCTSLSEILRHILASGADVTSANAKRYQKRGGFDTDDACMELRLSNFSLVKL 718
Db 383 EIPLTLHSVSELVRLCLRRSDVQEESEGSTD-----DNKDSAAFEDNEVQDEFLEKL 435
Qy 719 SSTSVYDLTPGKMKILHALCGKLLTLVSTRDFIEDYVDILRQAKQEPRELKAEQHRKER 778
Db 436 ETSEFFELTSEKQIQLTALCHRLMTYSVQD-----HMETR 472
Qy 779 EEAARIRKRKEELKEQEKQKKEQKELKEDQORNSTADISIGEEREDFDTIESKDT 838
Db 473 QQMSAELWKERLAVLKEENDKKRAEKQKKE-----MEAKNK 509
Qy 839 EQKELDQDMFTEDDDPGSHKRGRRGKRGQNGKFEFTQOE-QINCVTRELITADEEALK 897
Db 510 ENGVEN-----GLGKTRK-----KEIVKFEQVDTEADMDISAVKSRLL 551
Qy 898 QEHQKEKELLEK-IQSAIAC 917
Db 552 AIQAKKEREIOEREMKKGKISC 572

RESULT 34
ABB33805
ID ABB33805 standard; peptide; 572 AA.

AC ABB33805;

DT 04-FEB-2002 (first entry)

XX Peptide #1311 encoded by human foetal liver single exon probe.

XX Human; foetal liver; gene expression; single exon nucleic acid probe.

XX Homo sapiens.

XX WO200157277-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US000669.

XX 04-FEB-2000; 2000US-0180312P.

XX 26-MAY-2000; 2000US-0207456P.

XX 30-JUN-2000; 2000US-00608408.

XX 03-AUG-2000; 2000US-00632366.

XX 21-SEP-2000; 2000US-0234687P.

XX 27-SEP-2000; 2000US-0236359P.

XX 04-OCT-2000; 2000GB-00024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-483447/52.

XX Human genome-derived single exon nucleic acid probes useful for analyzing
XX gene expression in human foetal liver.

PS Claim 27; SEQ ID NO 26440; 639pp + Sequence Listing; English.
XX The invention relates to a single exon nucleic acid probe for measuring
CC human gene expression in a sample derived from human foetal liver. The
CC single exon nucleic acid probes may be used for predicting, measuring and
CC displaying gene expression in samples derived from human foetal liver. The
CC present sequence is a peptide encoded by a single exon nucleic acid probe
CC of the invention. Note: The sequence data for this patent did not form
CC part of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX

SQ Sequence 572 AA;

Query Match 3.7%; Score 318.5; DB 4; Length 572;

Best Local Similarity 24.8%; Pred. No. 5.2e-13;

Matches 124; Conservative 89; Mismatches 173; Indels 115; Gaps 20;

Qy 429 PPKRIHI-----SOEDNVANKQTLASYSRKATKRDKLLKQEEEMKSLAFKAKUKREKAD 483

Db 175 PPAALHLIAYYKKNDRKDSALSCVSKTA---RLSSDDRARLPPEELSLVQKRYE 230

Qy 484 ALEAKK-KEDDEKKREELKVIIEERLKKKEKERLKVREKERELREKKRYEYL 542

Db 231 LLEHKRWASMSERKEYLKK--KREELKK-LKAKAKERREKEMLERLEKQKRY---- 283

Qy 543 KQMSKPRDMCEDDLKELPEPTPVKT--RLPPEIFGDALMVLFLNAGFELFDLQDEPPD 600

Db 284 -----EDQELTG-KNLPAPRLVDTPEGLNTLFGDVAMVVEFLSCYSGLLLPDAQYP- 334

Qy 601 GVTLEVLBEALVGNDSGELCELLFFLTAIFQAIABE--BEEVAKEQLTDADTKGCSLK 658

Db 335 -ITAVSLMEAL-SADKGG-----FLYNRLVILLQTLLODEIAE---DYGELGMKLS 382

Qy 659 SLDDSDCTSLSEILRHILASGADVTSANAKRYQKRGGFDTDDACMELRLSNFSLVKL 718

Db 383 EIPLTLHSVSELVRLCLRRSDVQEESEGSTD-----DNKDSAAFEDNEVQDEFLEKL 435

Qy 719 SSTSVYDLTPGKMKILHALCGKLLTLVSTRDFIEDYVDILRQAKQEPRELKAEQHRKER 778

Db 436 ETSEFFELTSEKQIQLTALCHRLMTYSVQD-----HMETR 472

Qy 779 EEAARIRKRKEELKEQEKQKKEQKELKEDQORNSTADISIGEEREDFDTIESKDT 838

Db 473 QQMSAELWKERLAVLKEENDKKRAEKQKKE-----MEAKNK 509

Qy 839 EQKELDQDMFTEDDDPGSHKRGRRGKRGQNGKFEFTQOE-QINCVTRELITADEEALK 897

Db 510 ENGVEN-----GLGKTRK-----KEIVKFEQVDTEADMDISAVKSRLL 551

Qy 898 QEHQKEKELLEK-IQSAIAC 917

Db 552 AIQAKKEREIOEREMKKGKISC 572

RESULT 35
AAM27265

ID AAM27265 standard; protein; 572 AA.

AC AAM27265;

DT 17-OCT-2001 (first entry)

XX Peptide #1302 encoded by probe for measuring placental gene expression.

XX Probe; microarray; human; placenta; antenatal diagnosis;

XX genetic disorder.

XX Homo sapiens.

XX WO200157272-A2.

XX 09-AUG-2001.


```
PF 30-JAN-2001; 2001WO-US000663.
XX
PR 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-0060840B.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488897/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for analyzing
XX gene expression in human placenta.
XX
XX Claim 27; SEQ ID NO 27534; 654pp; English.
XX
XX The present invention relates to single exon nucleic acid probes (SENP:
XX see AA131315-AA157546). The present sequence is a peptide encoded by one
XX such probe. The probes are useful for producing a microarray for
XX predicting, measuring and displaying gene expression in samples derived
XX from human placenta. The probes are useful for antenatal diagnosis of
XX human genetic disorders
XX
XX Sequence 572 AA;
XX
Query Match 3.7%; Score 318.5; DB 4; Length 572;
Best Local Similarity 24.8%; Pred. No. 5.2e-13;
Matches 124; Conservative 89; Mismatches 173; Indels 115; Gaps 20;
Qy 429 PPKRIHI-----SQEDNVANKQTASYRSKATKRDKLLKQEMKSLAFKAKLKREKAD 483
Db 175 PPAALHLIAYYKENKDRKDSALSCVISTKA-----RLSSSDRARLPEELRSVQKRYE 230
Qy 484 ALEAKKK-EKEDKEKKBELKKIVEERLKKKEERLKKVEREKREKLREKRYVEYL 542
Db 231 LLEHKRWASMSERKEYLKK--KREELKKK-LKEKAKERREKEMLERLEKQRY-----283
Qy 543 KQWSKPRDMECDLDELPEPTPVKT--RLPPEIFGDALMVLFLNFAFGELFDLQDEFPD 600
Db 284 -----BDQELTG-KNLPAPRLVDTPEGLPNTLFGDVAMVVEFLSCYSGLLLPDAQYP- 334
Qy 601 GTVLEVLREALVGNDSGELCELLEFFLTAIFQATAE--BEVAKELQITADTKGCSLK 658
Db 335 -ITAVSLMEAL-SADKGG-----FLYNRLVILLQTLQDEIAE-----DYGELGMKLS 382
Qy 659 SLIDLSCVLSBTLRLHILASGADVTSANAKYRQKRGFGDATDDACMELRLSNPSLVKVL 718
Db 383 EIPTLHVSSELVRLCLRSDVQSESGD-----DNKDSAPFEDNEVQDEFLEKL 435
Qy 719 SSTSYDITLPEKMKILHALCKLTLVSTRDFIEDYVDILRQAKQEFRELKAEQHRKER 778
Db 436 ETSEFFELTSEKQLTALCHRLIMTVSVQD-----HMETR 472
Qy 779 EEMARIKREKELKEQEKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQK 838
Db 473 QQMSAELWKLAVLLEKNDKKAQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQ 509
Qy 839 EQKELQDMFTDEDDPGSHKGRGKRGQNGFKETROE-QINCVTRELLTADDEEALK 897
Db 510 ENKVEN-----GLGKTRK-----KEIVKFEQVDTAEADMSIAVKSRL 551
Qy 898 QEHQREKELLEK-IQSAIAC 917
Db 552 AIQAKKEREIQEREMKGIKSC 572
```

RESULT 36
ABB28622

```
ID ABB28622 standard; peptide; 572 AA.
XX
AC ABB28622;
XX
DT 01-FEB-2002 (first entry)
XX
DE Peptide #1273 encoded by breast cell single exon nucleic acid probe.
XX
KW Human; microarray; single exon probe; gene expression; breast; disease;
KW cancer.
XX
XX Homo sapiens.
XX
XX WO200157271-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US000662.
XX
XX 04-FEB-2000; 2000US-0180312P.
XX 26-MAY-2000; 2000US-0207456P.
XX 30-JUN-2000; 2000US-0060840B.
XX 03-AUG-2000; 2000US-00632366.
XX 21-SEP-2000; 2000US-0234687P.
XX 27-SEP-2000; 2000US-0236359P.
XX 04-OCT-2000; 2000GB-00024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-496933/54.
XX
XX New spatially-addressable set of single exon nucleic acid probes, useful
XX for measuring gene expression in sample derived from human breast,
XX comprises number of single exon nucleic acid probes.
XX
XX Claim 27; SEQ ID NO 11590; 327pp + Sequence Listing; English.
XX
XX The invention relates to a spatially-addressable set of single exon
XX nucleic acid probes for measuring gene expression in a sample derived
XX from human breast and Br 474 cells. The method involves contacting the
XX probes with a collection of detectably labelled nucleic acids derived
XX from mRNA of human breast, and then measuring the label bound to each
XX probe of the microarray. The probes are useful for verifying the
XX expression of regions of genomic DNA predicted to encode proteins. They
XX are useful for gene discovery, and for determining predisposition and/or
XX prognosing breast disease. Gene expression analysis is useful for
XX assessing the toxicity of chemical agents on cells. The microarray of
XX this invention presents a far greater diversity of probes for measuring
XX gene expression, with far less bias than expressed sequence tag
XX microarrays. The method is suitable for rapid production of functional
XX information from genomic sequence. The present sequence is a peptide
XX encoded by a single exon nucleic acid probe of the invention. Note: The
XX sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 572 AA;
XX
Query Match 3.7%; Score 318.5; DB 4; Length 572;
Best Local Similarity 24.8%; Pred. No. 5.2e-13;
Matches 124; Conservative 89; Mismatches 173; Indels 115; Gaps 20;
Qy 429 PPKRIHI-----SQEDNVANKQTASYRSKATKRDKLLKQEMKSLAFKAKLKREKAD 483
Db 175 PPAALHLIAYYKENKDRKDSALSCVISTKA-----RLSSSDRARLPEELRSVQKRYE 230
Qy 484 ALEAKKK-EKEDKEKKBELKKIVEERLKKKEERLKKVEREKREKLREKRYVEYL 542
Db 231 LLEHKRWASMSERKEYLKK--KREELKKK-LKEKAKERREKEMLERLEKQRY-----283
Qy 543 KQWSKPRDMECDLDELPEPTPVKT--RLPPEIFGDALMVLFLNFAFGELFDLQDEFPD 600
Db 284 -----BDQELTG-KNLPAPRLVDTPEGLPNTLFGDVAMVVEFLSCYSGLLLPDAQYP- 334
Qy 601 GTVLEVLREALVGNDSGELCELLEFFLTAIFQATAE--BEVAKELQITADTKGCSLK 658
Db 335 -ITAVSLMEAL-SADKGG-----FLYNRLVILLQTLQDEIAE-----DYGELGMKLS 382
Qy 659 SLIDLSCVLSBTLRLHILASGADVTSANAKYRQKRGFGDATDDACMELRLSNPSLVKVL 718
Db 383 EIPTLHVSSELVRLCLRSDVQSESGD-----DNKDSAPFEDNEVQDEFLEKL 435
Qy 719 SSTSYDITLPEKMKILHALCKLTLVSTRDFIEDYVDILRQAKQEFRELKAEQHRKER 778
Db 436 ETSEFFELTSEKQLTALCHRLIMTVSVQD-----HMETR 472
Qy 779 EEMARIKREKELKEQEKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQK 838
Db 473 QQMSAELWKLAVLLEKNDKKAQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQ 509
Qy 839 EQKELQDMFTDEDDPGSHKGRGKRGQNGFKETROE-QINCVTRELLTADDEEALK 897
Db 510 ENKVEN-----GLGKTRK-----KEIVKFEQVDTAEADMSIAVKSRL 551
Qy 898 QEHQREKELLEK-IQSAIAC 917
Db 552 AIQAKKEREIQEREMKGIKSC 572
```



```

PR 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
PA Penn SG, Hanzel DK, Chen W, Rank DR;
PI WPI; 2001-488900/53.
XX Human genome-derived single exon nucleic acid probes useful for analyzing
PT gene expression in human bone marrow.
PT gene expression in human bone marrow.
XX Example 4; SEQ ID NO 27284; 659pp + Sequence Listing; English.
XX The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC bone marrow. They can be used to measure gene expression in bone marrow
CC samples, which may enable the improved diagnosis and treatment of cancers
CC such as lymphoma, leukaemia and myeloma. The present sequence is a
CC protein encoded by one of the probes of the invention
XX
XX Sequence 572 AA;
SQ
Query Match 3.7%; Score 318.5; DB 4; Length 572;
Best Local Similarity 24.8%; Pred. No. 5.2e-13;
Matches 124; Conservative 89; Mismatches 173; Indels 115; Gaps 20;
QY 429 PPKRIHI-----SQEDNVANKOTLASYSRKATKRDKLLKOBEMKSLAFKAKLKREKAD 483
Db 175 PPAALHLIAYYKENDKREDKRSALSCVSKTA-----RLLSSEDRARLPEELRSVQKRYE 230
QY 484 ALEAKK-EKEDKEKKREELKIVVEERLKKKEERLKVREKERELREBKRYVEYL 542
Db 231 LLEHKRWASMSSEQRKEYLKK--KREELKKK-LKEKAKERREKEMLERLEKQKRY---- 283
QY 543 KQWSPREDMECDLDELPEPTPVKT--RLPEIFGDALMWLEFLNAGELFDLODEFPD 600
Db 284 -----EDQELTG-KNLPAPFLVDTPGGLPNTLFGDVAMVVFVFLSCYSGLLLPDAQYP- 334
QY 601 GVTLEVLBEALVGNDSGFLCELLFFLTALFQAIABE--BEEVAKEQLTDADTKGCSLK 658
Db 335 -ITAVSLMEAL-SADKGG-----FLYLNRLVILLQTLLODEIAE----DYGELGMKLS 382
QY 659 SLDLDSCTLSEILRLHILASGADVTSANAKRYQKRGGFDTDDACMELRLSNPSLVKKL 718
Db 383 EIPLTLHSVSELVRLCLRRSDVQEESESGDITD-----DNKDSAAFEDNEVDQEFLEKL 435
QY 719 SSTSVYDLTPGKMKILHALCGKLLTLVSTRDFIEDYVDILRQAKQEFRELKAEQHRKER 778
Db 436 ETSEFFELTSEBKQLTALCHRLIMLTYSVQD-----HMETR 472
QY 779 BEAARIRKKEEKLKEQEKMKKEQEKLEKEDQERNSTADISIGBEEREDPDTSTESKDT 838
Db 473 QQMSAELWKEKRLAVLKEENDKQAEKQKKE-----NEAKNK 509
QY 839 EOKELDQMFTEDEDDPGSHKRGKRGKQNGFKFTQOE-QINCVTRELTADBEELK 897
Db 510 ENKGVN-----GLGKTRK-----KEIVKFEPOVDTEADMISAVKSRLL 551
QY 898 QEHQKKEKLEK-IQSAIAC 917
Db 552 AIQAKKEREIOEREMKGIK 572
RESULT 39
AAM54571
ID AAM54571 standard; protein; 572 AA.
XX

```

```

AC AAM54571;
XX 05-NOV-2001 (first entry)
XX Human brain expressed single exon probe encoded protein SEQ ID NO: 26676.
XX Human; brain expressed exon; gene expression analysis; probe; microarray;
KW Alzheimer's disease; multiple sclerosis; schizophrenia; epilepsy; cancer.
XX Homo sapiens.
XX WO200157275-A2.
XX 09-AUG-2001.
XX 30-JAN-2001; 2001WO-US000667.
XX 04-FEB-2000; 2000US-0180312P.
XX 26-MAY-2000; 2000US-0207456P.
XX 30-JUN-2000; 2000US-00608408.
XX 03-AUG-2000; 2000US-00632366.
XX 21-SEP-2000; 2000US-0234687P.
XX 27-SEP-2000; 2000US-0236359P.
XX 04-OCT-2000; 2000GB-00024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-483446/52.
XX Single exon nucleic acid probes for analyzing gene expression in human
PT brains.
PT brains.
XX Example 4; SEQ ID NO 26676; 650pp + Sequence Listing; English.
XX The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC brain. They can be used to measure gene expression in brain cell samples,
CC which may enable the diagnosis and improved treatment of nervous system
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC epilepsy and cancers. The present sequence is a protein encoded by one of
CC the probes of the invention
XX
XX Sequence 572 AA;
SQ
Query Match 3.7%; Score 318.5; DB 4; Length 572;
Best Local Similarity 24.8%; Pred. No. 5.2e-13;
Matches 124; Conservative 89; Mismatches 173; Indels 115; Gaps 20;
QY 429 PPKRIHI-----SQEDNVANKOTLASYSRKATKRDKLLKOBEMKSLAFKAKLKREKAD 483
Db 175 PPAALHLIAYYKENDKREDKRSALSCVSKTA-----RLLSSEDRARLPEELRSVQKRYE 230
QY 484 ALEAKK-EKEDKEKKREELKIVVEERLKKKEERLKVREKERELREBKRYVEYL 542
Db 231 LLEHKRWASMSSEQRKEYLKK--KREELKKK-LKEKAKERREKEMLERLEKQKRY---- 283
QY 543 KQWSPREDMECDLDELPEPTPVKT--RLPEIFGDALMWLEFLNAGELFDLODEFPD 600
Db 284 -----EDQELTG-KNLPAPFLVDTPGGLPNTLFGDVAMVVFVFLSCYSGLLLPDAQYP- 334
QY 601 GVTLEVLBEALVGNDSGFLCELLFFLTALFQAIABE--BEEVAKEQLTDADTKGCSLK 658
Db 335 -ITAVSLMEAL-SADKGG-----FLYLNRLVILLQTLLODEIAE----DYGELGMKLS 382
QY 659 SLDLDSCTLSEILRLHILASGADVTSANAKRYQKRGGFDTDDACMELRLSNPSLVKKL 718
Db 383 EIPLTLHSVSELVRLCLRRSDVQEESESGDITD-----DNKDSAAFEDNEVDQEFLEKL 435
QY 719 SSTSVYDLTPGKMKILHALCGKLLTLVSTRDFIEDYVDILRQAKQEFRELKAEQHRKER 778
Db 436 ETSEFFELTSEBKQLTALCHRLIMLTYSVQD-----HMETR 472

```


Qy	806	KLKDEBORNSTADISIGEEREDPDTSTIESKDTQKELDQDMFTEDDDPGSHKRRGK	865
Db	1540	LLKKKEKEKEREKKEKEREQ-----KLKKEKEREKLAETEE-----RQREK	1584
Qy	866	RQNGKFKEFTROEQINCVTRELLTADEEALKQEHQR-----KEKE-LLEKTIQ	912
Db	1585	EREKLEKERAELKDLKEVKLEKEEQLEKKEKELKLEKKEKDKVKEKESLESBK	1644
Qy	913	SAIACTNIFPLGRDMYRRYWFIPSPGLFTIEDYSG-----TEDMLLRPS	960
Db	1645	LLISATVSNP-----WRRV-VEDTPPKLPFAVDYPSLGGKPKASPEKKRDEKLLPLGUT	1697
Qy	961	S-----FQNVQSDPQVSTKTGEPL-----MSESTNID-QGPRDHSVQLPK	1002
Db	1698	TPPKEVNDTFDFLSGLKPLEALPPLPALEPLEVKEDSKESVSLINFESPLQETASIR	1757
Qy	1003	PVHKPNRCFYSSCEQLDLQLEAL-----	1026
Db	1758	KISPPR-----GFTQ--NLILALCGSLHYENEQERIRETEAEVQPEVVTTPPEVYITLD	1811
Qy	1027	-----NSRGHRESALKETLQEKSRICQAOLARESEKPFHFDKQPQDS---KP	1071
Db	1812	TLALQKSTSTNNSGSEIIVMBEPPVKLSKHKRLKQLQLOQQOQOREREPDDELRP	1871
Qy	1072	TYSGRSSNAYD-----PSQCAEKQLLELRDLRDLFLDIEDRI	1108
Db	1872	LISIMCDSQLDVPDIPVRAMPDDDEGVEEPEPEQSLPEDLLHFGSSGVATNATTTD	1931
Qy	1109	YQGTFLGAIKVTDRHIWRSALESG-----RYELL-----SEENKENG	1145
Db	1932	SEGVPATTSDDNVYGSSTSNAPPHKLTKKLEHKINLIAAIEAATSSSTSAEESV	1991
Qy	1146	IKTNVEDVEIMEIDQTKVIKORLLGIKTETPTV-----STNASTPOS--VSS	1193
Db	1992	ETTVPDPTGLGLG-----IGLQSHSPGAAVAGSGSTTGGAATAASLPAAAGGVT	2040
Qy	1194	V-----VHVLAM-----ALFOIEQGIERRFLKAPL	1218
Db	2041	VGLDAGAPTAHALQOQQOQQOQPEQLPPPPHPPPEEPGPAIFQPNMLKQEEVEQL	2100
Qy	1219	DASDSGRSYKTVLDRWRSELSSASUSQVFLHLSTLDRSVIWSKS-----ILNARCKIC	1272
Db	2101	DGLALNN-----NNINNIIPTKONASPNPNTNPNAI	2135
Qy	1273	RKKGDAENMLDCDRG-HHTYICVRPKLTVPEDGWFCECRPK--QRCRRLSFRQPS	1329
Db	2136	PSSNKVQKAEVTDVADYPIRHSYGDHGLSTVWVNIIEVFDQLSMSSISTNSRLSLEGP	2195
Qy	1330	LES-----DEDVEDS-----MGEDDEVDGDEEGQSEEEYEV	1363
Db	2196	QSSPPLSLTNGMLMTNLMHFSESVTGTGVAQAQVHGNGANRFEVDDDEEBEEBELDE	2255
Qy	1364	EQEDDSQEEE	1374
Db	2256	EAEPPDDDERED	2266
RESULT 45			
ID	AA18320	standard; protein; 560 AA.	
XX	AA18320;		
AC	AA18320;		
XX	12-OCT-2001	(first entry)	
XX	Peptide #4754	encoded by probe for measuring cervical gene expression.	
DE	Probe; human; microarray;	gene expression; cervical epithelial cell;	
XX	cervical cancer.		
KW	Homo sapiens.		
XX			
OS			
XX			

Query Match 3.6%; Score 314; DB 4; Length 560;
Best Local Similarity 24.7%; Pred. No. 1e-12;

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 9, 2005, 13:54:12 ; Search time 26.8675 Seconds
(without alignments)
4242.638 Million cell updates/sec

Title: US-10-702-148-27
Perfect score: 7967
Sequence: 1 MAPLGRKPFPLVNPGLPGE.....NCFMMLVNTQFCMALTDVT 1527

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 65 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/1/aa/5A-COMB.pep:*
2: /cgn2_6/ptodata/1/aa/5B-COMB.pep:*
3: /cgn2_6/ptodata/1/aa/6A-COMB.pep:*
4: /cgn2_6/ptodata/1/aa/6B-COMB.pep:*
5: /cgn2_6/ptodata/1/aa/PCTUS-COMB.pep:*
6: /cgn2_6/ptodata/1/aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	7967	100.0	1527	4	US-09-418-710-27
2	7967	100.0	1527	4	US-09-839-478-27
3	7955	99.8	1531	4	US-09-418-710-29
4	7955	99.8	1531	4	US-09-839-478-29
5	7888	99.0	1525	4	US-09-418-710-69
6	7888	99.0	1525	4	US-09-839-479-68
7	7674	96.3	1540	4	US-09-949-016-7037
8	929.5	11.7	1674	4	US-09-418-710-1
9	929.5	11.7	1674	4	US-09-839-479-1
10	918	11.5	1673	4	US-09-418-710-70
11	918	11.5	1673	4	US-09-839-479-69
12	637	8.0	128	4	US-09-513-999C-7706
13	532	6.7	1972	4	US-09-418-710-21
14	532	6.7	1972	4	US-09-839-479-21
15	527.5	6.6	1876	4	US-09-418-710-71
16	527.5	6.6	1876	4	US-09-839-479-71
17	521.5	6.5	1878	4	US-09-418-710-13
18	521.5	6.5	1878	4	US-09-839-479-13
19	508.5	6.4	1969	4	US-09-418-710-72
20	508.5	6.4	1969	4	US-09-839-479-72
21	352	4.4	60	4	US-09-513-999C-7181
22	311.5	3.9	800	4	US-09-270-767-45282
23	304	3.8	2733	4	US-09-949-016-11433
24	303	3.8	3259	4	US-09-949-016-6507
25	300.5	3.8	1972	4	US-08-875-435B-3
26	299	3.8	1400	4	US-09-764-176-7
27	296	3.7	2375	4	US-09-538-092-1131

RESULT 1

US-09-418-710-27
; Sequence 27, Application US/09418710
; Patent No. 6596482
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042001
; CURRENT APPLICATION NUMBER: US/09/418,710
; CURRENT FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 27
; LENGTH: 1527
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-418-710-27

ALIGNMENTS

28	294.5	3.7	8991	4	US-08-714-741-32	Sequence 32, Appl
29	292.5	3.7	1857	4	US-09-917-254-91	Sequence 91, Appl
30	292.5	3.7	1972	4	US-09-538-092-1084	Sequence 1084, Ap
31	292.5	3.7	1984	4	US-09-949-016-7111	Sequence 7111, Ap
32	292.5	3.7	1984	4	US-09-949-016-7112	Sequence 7112, Ap
33	292.5	3.7	1984	4	US-09-949-016-7113	Sequence 7113, Ap
34	281.5	3.5	1972	4	US-08-875-435B-4	Sequence 4, Appl
35	281	3.5	65	4	US-09-418-710-68	Sequence 68, Appl
36	281	3.5	65	4	US-09-839-479-67	Sequence 67, Appl
37	281	3.5	2662	4	US-09-595-684B-31	Sequence 31, Appl
38	280.5	3.5	2663	4	US-09-538-092-1252	Sequence 1252, Ap
39	277	3.5	59	4	US-09-418-710-67	Sequence 67, Appl
40	277	3.5	59	4	US-09-839-479-66	Sequence 66, Appl
41	274.5	3.4	3924	4	US-09-538-092-1246	Sequence 1246, Ap
42	274	3.4	1898	2	US-08-056-200-94	Sequence 94, Appl
43	274	3.4	1898	2	US-08-800-644-94	Sequence 94, Appl
44	274	3.4	1898	4	US-09-538-092-1280	Sequence 1280, Ap
45	272	3.4	2753	4	US-09-949-016-7659	Sequence 7659, Ap
46	272	3.4	2753	4	US-09-949-016-7660	Sequence 7660, Ap
47	268	3.4	3210	4	US-09-538-092-1154	Sequence 1154, Ap
48	265	3.3	1976	4	US-09-538-092-1078	Sequence 1078, Ap
49	264	3.3	3248	1	US-08-353-700-1	Sequence 1, Appl
50	264	3.3	3248	5	PCT-US95-16216-1	Sequence 1, Appl
51	263	3.3	1805	1	US-07-853-913-2	Sequence 2, Appl
52	263	3.3	2482	1	US-08-328-254-6	Sequence 6, Appl
53	254	3.2	1939	3	US-09-310-187A-1	Sequence 1, Appl
54	254	3.2	1939	4	US-09-538-092-917	Sequence 917, App
55	254	3.2	1960	4	US-09-538-092-1077	Sequence 1077, Ap
56	253	3.2	1960	4	US-09-949-016-10872	Sequence 10872, A
57	252.5	3.2	1596	4	US-08-978-277A-4	Sequence 4, Appl
58	252	3.2	2101	1	US-08-466-390-4	Sequence 4, Appl
59	252	3.2	2101	1	US-08-470-950-4	Sequence 4, Appl
60	252	3.2	2101	1	US-08-467-781-4	Sequence 4, Appl
61	252	3.2	2101	2	US-08-483-924-4	Sequence 4, Appl
62	252	3.2	2101	3	US-09-452-294-1	Sequence 1, Appl
63	251.5	3.2	1939	4	US-09-538-092-915	Sequence 915, App
64	251.5	3.2	1939	4	US-09-949-016-11104	Sequence 11104, A
65	251.5	3.2	1959	4	US-09-949-016-8134	Sequence 8134, Ap

Query Match 100.0%; Score 7967; DB 4; Length 1527;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1527; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MAPLGRKPFPLVNPGLPGEPPFTTPTQEAFTREEYEARLYSERIWTCKTSSQL 60
|||||

Db 1 MAPLGRKPPFLVNPVLPGBEPFTTIPHTQEAFTREEYEARLERYSERIWTCKSGSSQL 60
Qy 61 THKEAWBEEQVVAELLKEEFPNAYEKLVLVEMVHNTASLEKLVDTAWLEIMTKYAVGEEC 120
Db 61 THKEAWBEEQVVAELLKEEFPNAYEKLVLVEMVHNTASLEKLVDTAWLEIMTKYAVGEEC 120
Qy 121 DFEVCKEKMVKVIKVIHPLKVDDEATEKSDGACDSSPSKENSQIAODHOKKETV 180
Db 121 DFEVCKEKMVKVIKVIHPLKVDDEATEKSDGACDSSPSKENSQIAODHOKKETV 180
Qy 181 KEDEGRRESINDRARRSPKLTSLKGGKRWAPKFLPHKYDVVKLQNEDKIISNVPADS 240
Db 181 KEDEGRRESINDRARRSPKLTSLKGGKRWAPKFLPHKYDVVKLQNEDKIISNVPADS 240
Qy 241 LIRTERPNKEIVRPIRHNAIRAGTGNAPWVDEDELVKKYSLSKPSDFSFLDDPKYMT 300
Db 241 LIRTERPNKEIVRPIRHNAIRAGTGNAPWVDEDELVKKYSLSKPSDFSFLDDPKYMT 300
Qy 301 LNPSTKRKNTGSPDRKPKSKTDNSLSPLNPKLWCHVHLKLSLGSPLKVKNSKNSK 360
Db 301 LNPSTKRKNTGSPDRKPKSKTDNSLSPLNPKLWCHVHLKLSLGSPLKVKNSKNSK 360
Qy 361 SPEEHLKEEMKMSPNKLTNPHIPKGGPPAKPKGKSDKPLKAKGRSGKILNGQKSTGN 420
Db 361 SPEEHLKEEMKMSPNKLTNPHIPKGGPPAKPKGKSDKPLKAKGRSGKILNGQKSTGN 420
Qy 421 SKSPKKGKLTPTKTKMKTLLDMAKGTQMTAPRNSGGTPTRTSKPKHKLPPAALHLIA 480
Db 421 SKSPKKGKLTPTKTKMKTLLDMAKGTQMTAPRNSGGTPTRTSKPKHKLPPAALHLIA 480
Qy 481 YKKNKREDKESALSVCISTARLLSGEDRARLPEELRSLVOKYELLEHKKRWASWE 540
Db 481 YKKNKREDKESALSVCISTARLLSGEDRARLPEELRSLVOKYELLEHKKRWASWE 540
Qy 541 EORKEYLKKREELKKLKEKAKERREKEMLEKQKRYEDELTKGNLPAFLVDTPE 600
Db 541 EORKEYLKKREELKKLKEKAKERREKEMLEKQKRYEDELTKGNLPAFLVDTPE 600
Qy 601 GLPNTLPFGDVAMVVEFLSCYGLLLPDAQYPTITAVSLMEALSADKGGFFLYLNRLVILLQ 660
Db 601 GLPNTLPFGDVAMVVEFLSCYGLLLPDAQYPTITAVSLMEALSADKGGFFLYLNRLVILLQ 660
Qy 661 TLLQDEIAEDYCELGKMLSEIPLTLHVSSELVRLCLRRSDVOESEGSDTDNDKDSAAFE 720
Db 661 TLLQDEIAEDYCELGKMLSEIPLTLHVSSELVRLCLRRSDVOESEGSDTDNDKDSAAFE 720
Qy 721 DNEVQDEFLEKLETSEFFELTSEEKLIQILTALCHRLMTYSVQDHMETRQMSAELWKER 780
Db 721 DNEVQDEFLEKLETSEFFELTSEEKLIQILTALCHRLMTYSVQDHMETRQMSAELWKER 780
Qy 781 LAVLKEENDKKAEBKQKREKEMAKNKGKENGKGLGKTDRKRIIVKFPQVDTEADMIS 840
Db 781 LAVLKEENDKKAEBKQKREKEMAKNKGKENGKGLGKTDRKRIIVKFPQVDTEADMIS 840
Qy 841 AVKSRLLAIQAKKRETOEREMVKLEQAESEERIRKHAAAEKAFQEGIAKAKLVMR 900
Db 841 AVKSRLLAIQAKKRETOEREMVKLEQAESEERIRKHAAAEKAFQEGIAKAKLVMR 900
Qy 901 TPIGTDRNHRNRYLFSDEVPGLFIEKGWVHDSIDYRFNHHCKDHTVSGDEDYCPRSKKN 960
Db 901 TPIGTDRNHRNRYLFSDEVPGLFIEKGWVHDSIDYRFNHHCKDHTVSGDEDYCPRSKKN 960
Qy 961 LQKNASMTQHGTATAVETTPKQGNLWFLCDSQKELDELNCLHPQGIRESQKLER 1020
Db 961 LQKNASMTQHGTATAVETTPKQGNLWFLCDSQKELDELNCLHPQGIRESQKLER 1020
Qy 1021 LEKRVQDIIHSHLARKPNLGLKSCDGOELLNFRSLDI EVA TRLQKGLGVSETSEF 1080
Db 1021 LEKRVQDIIHSHLARKPNLGLKSCDGOELLNFRSLDI EVA TRLQKGLGVSETSEF 1080
Qy 1081 EARVISLEKLDKDFGECVIALQASVKIKFLOQGFMAPKQKRKLQSDSDSAKTEVDEBKQV 1140
Db 1081 EARVISLEKLDKDFGECVIALQASVKIKFLOQGFMAPKQKRKLQSDSDSAKTEVDEBKQV 1140

Qy 1141 BEAKVASALEKWKTAIRAEQTFSRMHVLLGMLDACIKWDMSAENARCKVCPPKGGEDKLI 1200
Db 1141 BEAKVASALEKWKTAIRAEQTFSRMHVLLGMLDACIKWDMSAENARCKVCPPKGGEDKLI 1200
Qy 1201 LCDECNKAFHLFCLRLPALYEPDGEWQCPACOPATARRNSRGRNNTESASDESDES 1260
Db 1201 LCDECNKAFHLFCLRLPALYEPDGEWQCPACOPATARRNSRGRNNTESASDESDES 1260
Qy 1261 EEEEEEEBEEDEYEVAGLRPRKRTIRGKHSVIPPAAARSGRRPGKXPHSTRSQPKAP 1320
Db 1261 EEEEEEEBEEDEYEVAGLRPRKRTIRGKHSVIPPAAARSGRRPGKXPHSTRSQPKAP 1320
Qy 1321 PVDDAEVDDELVLQTKRSSRROSLELOKCEEILHLKIVKRFSPFPPEVTRDEAEDYDVI 1380
Db 1321 PVDDAEVDDELVLQTKRSSRROSLELOKCEEILHLKIVKRFSPFPPEVTRDEAEDYDVI 1380
Qy 1381 THPMDFTQVQNKSCGYSYRSVQEFITDMKQVFTNAEVNCRGSHVLSQVKTQECLVILL 1440
Db 1381 THPMDFTQVQNKSCGYSYRSVQEFITDMKQVFTNAEVNCRGSHVLSQVKTQECLVILL 1440
Qy 1441 HKHLPCHPYVRRKRRKFPDLAEDBGDSEPEBAVGOSRDEDRRSRBAEQEWLQDTSLSYA 1500
Db 1441 HKHLPCHPYVRRKRRKFPDLAEDBGDSEPEBAVGOSRDEDRRSRBAEQEWLQDTSLSYA 1500
Qy 1501 KINSKDHNCFMMLVNTQFCMALTDVT 1527
Db 1501 KINSKDHNCFMMLVNTQFCMALTDVT 1527

RESULT 2

US-09-839-479-27
; Sequence 27, Application US/09839479
; Patent No. 6727222
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTONAL REGULATOR
; FILE REFERENCE: 06501-042002
; CURRENT APPLICATION NUMBER: US/09/839,479
; CURRENT FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 27
; LENGTH: 1527
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-839-479-27

Query Match 100.0%; Score 7967; DB 4; Length 1527;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1527; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MAPLGRKPPFLVNPVLPGBEPFTTIPHTQEAFTREEYEARLERYSERIWTCKSGSSQL 60
Db 1 MAPLGRKPPFLVNPVLPGBEPFTTIPHTQEAFTREEYEARLERYSERIWTCKSGSSQL 60
Qy 61 THKEAWBEEQVVAELLKEEFPNAYEKLVLVEMVHNTASLEKLVDTAWLEIMTKYAVGEEC 120
Db 61 THKEAWBEEQVVAELLKEEFPNAYEKLVLVEMVHNTASLEKLVDTAWLEIMTKYAVGEEC 120
Qy 121 DFEVCKEKMVKVIKVIHPLKVDDEATEKSDGACDSSPSKENSQIAODHOKKETV 180
Db 121 DFEVCKEKMVKVIKVIHPLKVDDEATEKSDGACDSSPSKENSQIAODHOKKETV 180
Qy 181 KEDEGRRESINDRARRSPKLTSLKGGKRWAPKFLPHKYDVVKLQNEDKIISNVPADS 240

Db 181 ||||| KEDEGRESINDRARRSPKLP|TSLKGERKWA|PKFLPHKYDVKLQNEDKII|SNVPADS 240
Qy 241 ||||| LIRTERPNKEI|VRYE|FIRHNALRAGT|GENAPWVE|DELVKYSLPKS|FSDFLDDPYKYMT 300
Db 241 ||||| LIRTERPNKEI|VRYE|FIRHNALRAGT|GENAPWVE|DELVKYSLPKS|FSDFLDDPYKYMT 300
Qy 301 ||||| LNPSTRKNTGSDPRKPS|KSKT|DNS|SLSP|NPKLWCHVHLK|KSLG|SP|LKYNSK|NSK 360
Db 301 ||||| LNPSTRKNTGSDPRKPS|KSKT|DNS|SLSP|NPKLWCHVHLK|KSLG|SP|LKYNSK|NSK 360
Qy 361 ||||| SPEEHLEEMKMS|PNK|L|TN|FHI|PKK|GPPAK|PGH|SDK|PLKAK|GRS|KG|ILNG|QK|STGN 420
Db 361 ||||| SPEEHLEEMKMS|PNK|L|TN|FHI|PKK|GPPAK|PGH|SDK|PLKAK|GRS|KG|ILNG|QK|STGN 420
Qy 421 ||||| SKSPK|GL|T|PK|TK|MQ|MT|L|L|MAK|GT|QK|MT|APR|NS|GGT|P|RT|SS|K|PH|KL|PP|AA|L|H|L|A 480
Db 421 ||||| SKSPK|GL|T|PK|TK|MQ|MT|L|L|MAK|GT|QK|MT|APR|NS|GGT|P|RT|SS|K|PH|KL|PP|AA|L|H|L|A 480
Qy 481 ||||| YKENK|D|RED|K|S|A|L|S|C|V|I|S|K|T|A|L|L|S|S|E|D|R|A|L|P|E|L|R|S|L|V|Q|R|Y|E|L|L|E|H|K|K|R|W|A|S|G|E 540
Db 481 ||||| YKENK|D|RED|K|S|A|L|S|C|V|I|S|K|T|A|L|L|S|S|E|D|R|A|L|P|E|L|R|S|L|V|Q|R|Y|E|L|L|E|H|K|K|R|W|A|S|G|E 540
Qy 541 ||||| EQRKE|Y|L|K|K|B|E|L|K|K|K|E|K|A|K|E|R|E|K|E|W|L|R|E|K|Y|E|D|E|L|T|G|K|N|L|P|A|R|L|V|D|T|P|E 600
Db 541 ||||| EQRKE|Y|L|K|K|B|E|L|K|K|K|E|K|A|K|E|R|E|K|E|W|L|R|E|K|Y|E|D|E|L|T|G|K|N|L|P|A|R|L|V|D|T|P|E 600
Qy 601 ||||| GLPNT|L|FGD|VAM|V|E|L|S|C|Y|G|L|L|L|P|D|A|O|Y|P|I|T|A|V|S|L|ME|A|L|S|A|D|K|G|F|L|Y|L|N|R|V|L|L|Q 660
Db 601 ||||| GLPNT|L|FGD|VAM|V|E|L|S|C|Y|G|L|L|L|P|D|A|O|Y|P|I|T|A|V|S|L|ME|A|L|S|A|D|K|G|F|L|Y|L|N|R|V|L|L|Q 660
Qy 661 ||||| TL|Q|B|I|A|E|D|Y|G|E|L|G|K|L|S|E|I|P|U|L|H|S|V|S|E|L|V|R|L|C|R|R|S|D|V|Q|E|S|G|S|T|D|D|K|O|S|A|A|F|E 720
Db 661 ||||| TL|Q|B|I|A|E|D|Y|G|E|L|G|K|L|S|E|I|P|U|L|H|S|V|S|E|L|V|R|L|C|R|R|S|D|V|Q|E|S|G|S|T|D|D|K|O|S|A|A|F|E 720
Qy 721 ||||| DNEVQ|D|E|F|E|K|E|L|E|T|S|F|F|E|L|T|S|E|K|Q|I|L|T|A|L|C|R|I|L|T|Y|S|V|Q|D|H|E|T|R|Q|Q|S|A|E|L|W|K|R 780
Db 721 ||||| DNEVQ|D|E|F|E|K|E|L|E|T|S|F|F|E|L|T|S|E|K|Q|I|L|T|A|L|C|R|I|L|T|Y|S|V|Q|D|H|E|T|R|Q|Q|S|A|E|L|W|K|R 780
Qy 781 ||||| L|A|V|L|K|E|N|D|K|K|A|E|K|K|R|K|E|K|A|K|N|K|E|N|G|K|V|E|N|G|L|K|T|D|R|K|R|I|V|K|F|E|P|Q|V|D|T|E|A|R|D|M|L|S 840
Db 781 ||||| L|A|V|L|K|E|N|D|K|K|A|E|K|K|R|K|E|K|A|K|N|K|E|N|G|K|V|E|N|G|L|K|T|D|R|K|R|I|V|K|F|E|P|Q|V|D|T|E|A|R|D|M|L|S 840
Qy 841 ||||| A|V|K|R|R|L|A|I|O|A|K|K|R|E|I|O|E|R|M|K|V|K|L|R|Q|A|E|E|R|I|R|K|H|K|A|A|E|K|A|F|Q|B|G|I|A|K|A|K|L|V|R 900
Db 841 ||||| A|V|K|R|R|L|A|I|O|A|K|K|R|E|I|O|E|R|M|K|V|K|L|R|Q|A|E|E|R|I|R|K|H|K|A|A|E|K|A|F|Q|B|G|I|A|K|A|K|L|V|R 900
Qy 901 ||||| T|P|I|G|T|R|N|R|N|Y|L|F|S|D|E|V|P|G|L|F|I|E|K|G|W|H|D|S|I|D|Y|R|F|N|H|C|K|D|H|T|V|S|G|D|E|D|Y|C|P|R|S|K|K|A|N 960
Db 901 ||||| T|P|I|G|T|R|N|R|N|Y|L|F|S|D|E|V|P|G|L|F|I|E|K|G|W|H|D|S|I|D|Y|R|F|N|H|C|K|D|H|T|V|S|G|D|E|D|Y|C|P|R|S|K|K|A|N 960
Qy 961 ||||| L|G|K|N|A|S|M|N|T|O|H|G|T|A|T|E|V|A|V|E|T|T|P|K|Q|Q|N|L|W|F|L|C|S|O|K|E|L|D|E|L|N|C|L|H|P|Q|G|R|E|S|Q|L|K|R 1020
Db 961 ||||| L|G|K|N|A|S|M|N|T|O|H|G|T|A|T|E|V|A|V|E|T|T|P|K|Q|Q|N|L|W|F|L|C|S|O|K|E|L|D|E|L|N|C|L|H|P|Q|G|R|E|S|Q|L|K|R 1020
Qy 1021 ||||| L|E|K|R|Y|O|D|I|I|H|S|I|H|L|A|R|K|P|N|L|G|K|S|C|D|G|N|Q|E|L|N|F|L|R|S|D|I|E|V|A|R|L|Q|K|G|L|G|V|E|T|S|E|F 1080
Db 1021 ||||| L|E|K|R|Y|O|D|I|I|H|S|I|H|L|A|R|K|P|N|L|G|K|S|C|D|G|N|Q|E|L|N|F|L|R|S|D|I|E|V|A|R|L|Q|K|G|L|G|V|E|T|S|E|F 1080
Qy 1081 ||||| E|A|R|V|S|L|E|K|L|K|D|F|G|E|C|V|I|A|O|A|S|V|I|K|K|F|L|Q|G|F|A|P|K|Q|R|K|L|Q|S|E|D|S|A|K|T|E|V|D|E|K|Q|V 1140
Db 1081 ||||| E|A|R|V|S|L|E|K|L|K|D|F|G|E|C|V|I|A|O|A|S|V|I|K|K|F|L|Q|G|F|A|P|K|Q|R|K|L|Q|S|E|D|S|A|K|T|E|V|D|E|K|Q|V 1140
Qy 1141 ||||| E|A|K|V|A|S|A|L|E|K|W|K|T|A|I|R|E|A|Q|T|S|R|H|V|L|G|L|G|L|D|A|C|I|K|W|D|S|A|E|N|A|R|C|K|V|C|P|K|G|G|D|D|K|L| 1200
Db 1141 ||||| E|A|K|V|A|S|A|L|E|K|W|K|T|A|I|R|E|A|Q|T|S|R|H|V|L|G|L|G|L|D|A|C|I|K|W|D|S|A|E|N|A|R|C|K|V|C|P|K|G|G|D|D|K|L| 1200
Qy 1201 ||||| L|C|D|E|C|N|K|A|F|H|L|F|C|L|R|A|L|Y|E|V|D|G|E|W|Q|C|P|A|C|O|P|A|R|R|N|R|S|R|G|R|N|Y|T|E|S|A|S|D|S|E|D|S| 1260
Db 1201 ||||| L|C|D|E|C|N|K|A|F|H|L|F|C|L|R|A|L|Y|E|V|D|G|E|W|Q|C|P|A|C|O|P|A|R|R|N|R|S|R|G|R|N|Y|T|E|S|A|S|D|S|E|D|S| 1260
Qy 1261 ||||| E|E|E|E|E|E|E|E|E|E|D|E|V|A|G|R|L|R|P|R|K|T|I|R|G|K|H|S|V|I|P|P|A|R|S|G|R|R|P|G|K|P|H|S|R|T|R|S|Q|P|K|A|P 1320

Db 1261 ||||| E|E|E|E|E|E|E|E|E|E|D|E|V|A|G|R|L|R|P|R|K|T|I|R|G|K|H|S|V|I|P|P|A|R|S|G|R|R|P|G|K|P|H|S|R|T|R|S|Q|P|K|A|P 1320
Qy 1321 ||||| P|V|D|A|E|V|D|E|L|V|L|Q|T|K|R|S|R|R|Q|S|L|E|Q|C|E|I|L|H|K|I|V|K|R|F|S|W|P|R|P|V|T|R|D|E|A|E|D|Y|D|V|I 1380
Db 1321 ||||| P|V|D|A|E|V|D|E|L|V|L|Q|T|K|R|S|R|R|Q|S|L|E|Q|C|E|I|L|H|K|I|V|K|R|F|S|W|P|R|P|V|T|R|D|E|A|E|D|Y|D|V|I 1380
Qy 1381 ||||| T|H|P|M|D|F|Q|T|V|O|N|K|C|S|G|Y|S|V|Q|E|F|L|T|D|M|K|Q|V|T|F|N|A|E|V|Y|N|R|C|G|S|H|V|L|S|C|M|V|K|T|E|Q|C|L|V|L|L 1440
Db 1381 ||||| T|H|P|M|D|F|Q|T|V|O|N|K|C|S|G|Y|S|V|Q|E|F|L|T|D|M|K|Q|V|T|F|N|A|E|V|Y|N|R|C|G|S|H|V|L|S|C|M|V|K|T|E|Q|C|L|V|L|L 1440
Qy 1441 ||||| H|G|H|L|D|C|H|P|V|R|R|K|K|F|P|D|R|A|E|D|G|D|S|E|P|A|V|Q|S|R|D|E|R|R|S|R|E|A|E|I|O|E|W|L|O|D|T|S|L|Y|S|A 1500
Db 1441 ||||| H|K|H|L|P|G|H|P|V|R|R|K|K|F|P|D|R|A|E|D|G|D|S|E|P|A|V|Q|S|R|D|E|R|R|S|R|E|A|E|I|O|E|W|L|O|D|T|S|L|Y|S|A 1500
Qy 1501 ||||| K|I|N|S|K|D|H|N|C|F|M|L|V|N|T|Q|F|C|M|A|L|T|D|T|V|T 1527
Db 1501 ||||| K|I|N|S|K|D|H|N|C|F|M|L|V|N|T|Q|F|C|M|A|L|T|D|T|V|T 1527
RESULT 3
US-09-418-710-29
; Sequence 29, Application US/09418710
; Patent No. 6596482
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042001
; CURRENT APPLICATION NUMBER: US/09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 29
; LENGTH: 1531
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-418-710-29
Query Match 99.8%; Score 7955; DB 4; Length 1531;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 1527; Conservative 0; Mismatches 0; Indels 4; Gaps 1;
Qy 1 MAPLGRKPPFLVNPPLPGSEPPFTTIPHTQEAFTREYEYRLEARYSERIWTCKSTGSSQL 60
Db 1 MAPLGRKPPFLVNPPLPGSEPPFTTIPHTQEAFTREYEYRLEARYSERIWTCKSTGSSQL 60
Qy 61 THKEAWEBEQEAELLKEEFPAPWYKLVLEMYHHTASLEKLVDTAWLEIMTKYAVGEEC 120
Db 61 THKEAWEBEQEAELLKEEFPAPWYKLVLEMYHHTASLEKLVDTAWLEIMTKYAVGEEC 120
Qy 121 DPEVGEKMKLVKI|VKI|H|P|L|E|K|V|D|E|A|E|T|K|S|D|G|A|C|D|S|P|S|D|K|E|N|S|S|Q|A|Q|H|Q|K|E|T|V 180
Db 121 DPEVGEKMKLVKI|VKI|H|P|L|E|K|V|D|E|A|E|T|K|S|D|G|A|C|D|S|P|S|D|K|E|N|S|S|Q|A|Q|H|Q|K|E|T|V 180
Qy 181 KEDEGRRESINDRARRSPKLP|TSLKGERKWA|PKFLPHKYDVKLQNEDKII|SNVPADS 240
Db 181 KEDEGRRESINDRARRSPKLP|TSLKGERKWA|PKFLPHKYDVKLQNEDKII|SNVPADS 240
Qy 241 LIRTERPNKEI|VRYE|FIRHNALRAGT|GENAPWVE|DELVKYSLPKS|FSDFLDDPYKYMT 300
Db 241 LIRTERPNKEI|VRYE|FIRHNALRAGT|GENAPWVE|DELVKYSLPKS|FSDFLDDPYKYMT 300
Qy 301 LNPSTRKNTGSDPRKPS|KSKT|DNS|SLSP|NPKLWCHVHLK|KSLG|SP|LKYNSK|NSK 360
Db 301 LNPSTRKNTGSDPRKPS|KSKT|DNS|SLSP|NPKLWCHVHLK|KSLG|SP|LKYNSK|NSK 360
Qy 361 SPEEHLEEMKMS|PNK|L|TN|FHI|PKK|GPPAK|PGH|SDK|PLKAK|GRS|KG|ILNG|QK|STGN 420

```
Db 361 SPEEHLEEMKWMKSNKLTHTNPHIPKGGPPAKPKGKSHDKPLKAKGRSKGILNGQKSTGN 420
Qy 421 SKSPKGLKTPKTKWKQMTLLDMAKGTQKMTAPRNSGGTPTRTSKPHKHLPPAALHLIA 480
Db 421 SKSPKGLKTPKTKWKQMTLLDMAKGTQKMTAPRNSGGTPTRTSKPHKHLPPAALHLIA 480
Qy 481 YYKENKREDKESALSVCISKTARLLSSDDRARLPEELRSLVOKYELLEHKKRWASKE 540
Db 481 YYKENKREDKESALSVCISKTARLLSSDDRARLPEELRSLVOKYELLEHKKRWASKE 540
Qy 541 EQRKEYLKKRBEELKKLKEKAKERREKEMLEERLEKQKRYEYDEQELTGKNLPAFRLVDTPE 600
Db 541 EQRKEYLKKRBEELKKLKEKAKERREKEMLEERLEKQKRYEYDEQELTGKNLPAFRLVDTPE 600
Qy 601 GLPNTLFGDVAMVVEFLSCYSGILLPDAQYPTAVSLMEALSADKGGFLYLNRLVLI 657
Db 601 GLPNTLFGDVAMVVEFLSCYSGILLPDAQYPTAVSLMEALSADKGGFLYLNRLVLI 660
Qy 658 -LLQTLLODEIAEDYGEGLMKLSEIPLTLHSHVSELVRLCLRSDVQBESEGSDDTDNKS 716
Db 661 TLQTLLODEIAEDYGEGLMKLSEIPLTLHSHVSELVRLCLRSDVQBESEGSDDTDNKS 720
Qy 717 AAFEDNEVQDEPLEKLETSFEPFELTSEELQILTALCHRLMTYSVDHMETRQMSAEL 776
Db 721 AAFEDNEVQDEPLEKLETSFEPFELTSEELQILTALCHRLMTYSVDHMETRQMSAEL 780
Qy 777 WKERLAVLKEENDKKRAEKQKREKEMAKNKGKVENGLGKTDKRRIVKPEPQVDTAE 836
Db 781 WKERLAVLKEENDKKRAEKQKREKEMAKNKGKVENGLGKTDKRRIVKPEPQVDTAE 840
Qy 837 DMISAVKSRLLAIQAKKERETOEREMKVKLROAEERIRKHAAAEKAFQEGTAKAL 896
Db 841 DMISAVKSRLLAIQAKKERETOEREMKVKLROAEERIRKHAAAEKAFQEGTAKAL 900
Qy 897 VMRRTPIGTDNRHNYWLPFSDEVPGLFEKGWVHDSIDYRFNHHCXKDHVTSDEDCPRS 956
Db 901 VMRRTPIGTDNRHNYWLPFSDEVPGLFEKGWVHDSIDYRFNHHCXKDHVTSDEDCPRS 960
Qy 957 KXANLGKNAAMNTQHGTA TEVA VETTPKQGNLWFLCDSQKELDELLNCLHPQIGRESQ 1016
Db 961 KXANLGKNAAMNTQHGTA TEVA VETTPKQGNLWFLCDSQKELDELLNCLHPQIGRESQ 1020
Qy 1017 LKERLEKRYQDIIHSHILARKNGLKSCDGNQELLNLFSLIEVATRLQKGGLYVEE 1076
Db 1021 LKERLEKRYQDIIHSHILARKNGLKSCDGNQELLNLFSLIEVATRLQKGGLYVEE 1080
Qy 1077 TSEFEARVLSLEKLDGFCGCVIALQASVTKKFLQGFMAPKQKRRKLQSDSAKTEBEVD 1136
Db 1081 TSEFEARVLSLEKLDGFCGCVIALQASVTKKFLQGFMAPKQKRRKLQSDSAKTEBEVD 1140
Qy 1137 KQWVEAKVASALEKWKTAIRBAQTFSRMHVLLGLMDACIKWMSAENARCKVCPKKGED 1196
Db 1141 KQWVEAKVASALEKWKTAIRBAQTFSRMHVLLGLMDACIKWMSAENARCKVCPKKGED 1200
Qy 1197 DKLLCDECNKAFHLFCRLPALYEPDGEWQCPACOPATARNRGRNVTESASEDS 1256
Db 1201 DKLLCDECNKAFHLFCRLPALYEPDGEWQCPACOPATARNRGRNVTESASEDS 1260
Qy 1257 DESDEEEEBEEDYEVAGLRPRKTTIRGHSHVTPPAARSGRRPKKPHSTRSQ 1316
Db 1261 DESDEEEEBEEDYEVAGLRPRKTTIRGHSHVTPPAARSGRRPKKPHSTRSQ 1320
Qy 1317 PKAPPVDDAEVDDELVTQKRSRRSLELQKEEELHKTIVKRFPSWFPFEPVTRDEADY 1376
Db 1321 PKAPPVDDAEVDDELVTQKRSRRSLELQKEEELHKTIVKRFPSWFPFEPVTRDEADY 1380
Qy 1377 YDVI THPMDFOVQNKSCGYSYRQVEFLTDMKQFTNAEVNCRGSHVLSCMWTEOCL 1436
Db 1381 YDVI THPMDFOVQNKSCGYSYRQVEFLTDMKQFTNAEVNCRGSHVLSCMWTEOCL 1440
Qy 1437 VVLLHKLHPGHPYVRKRRKKFPDRLAEDGSDSEPAVGQSRDEDRRSREAEIQEWLQDTS 1496
Db 1441 VVLLHKLHPGHPYVRKRRKKFPDRLAEDGSDSEPAVGQSRDEDRRSREAEIQEWLQDTS 1500
```

```
Qy 1497 LYSAKINSKDHNCFMMLVNTQFCMALTDVT 1527
Db 1501 LYSAKINSKDHNCFMMLVNTQFCMALTDVT 1531
```

RESULT 4

```
US-09-839-479-29
; Sequence 29, Application US/09839479
; Patent No. 6727222
```

GENERAL INFORMATION:

```
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042002
; CURRENT APPLICATION NUMBER: US/09/839,479
; CURRENT FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 09/418,710
; FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 29
; LENGTH: 1531
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-839-479-29
```

```
Query Match 99.8%; Score 7955; DB 4; Length 1531;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 1527; Conservative 0; Mismatches 0; Indels 4; Gaps 1;
```

```
Qy 1 MAPLGRKPPFLVNPPLPGEEPFTTIPHTQEAFTREEYEARLERYSERIWTCKSGSSQL 60
Db 1 MAPLGRKPPFLVNPPLPGEEPFTTIPHTQEAFTREEYEARLERYSERIWTCKSGSSQL 60
Qy 61 THKEAWEEQVEAEELLKEEFPWYKLVLEMVHHTASLEKLVDTAWLEIMTKYAVGEEC 120
Db 61 THKEAWEEQVEAEELLKEEFPWYKLVLEMVHHTASLEKLVDTAWLEIMTKYAVGEEC 120
Qy 121 DPEVGEKQMLKVKIWIHPLEKVDDEATEKSDGACDSPSSDKENSSQIAQDHQKETT 180
Db 121 DPEVGEKQMLKVKIWIHPLEKVDDEATEKSDGACDSPSSDKENSSQIAQDHQKETT 180
Qy 181 KEDEGRRESINDRARRSRPKLPTSLKGERKWAPPKFLPHKYVDVKLQNEDKIISNVPADS 240
Db 181 KEDEGRRESINDRARRSRPKLPTSLKGERKWAPPKFLPHKYVDVKLQNEDKIISNVPADS 240
Qy 241 LIRTERPPNKEIVRYFIRHNALRAGTGENAPWVVEDELVKYKYSLPSPKSFDFLLDPKYMT 300
Db 241 LIRTERPPNKEIVRYFIRHNALRAGTGENAPWVVEDELVKYKYSLPSPKSFDFLLDPKYMT 300
Qy 301 LNPSTKRNKTSQSDPKPSKSKTDNSSLSPNLPKLVCHVHLKKSLSGSLPKVKNKSK 360
Db 301 LNPSTKRNKTSQSDPKPSKSKTDNSSLSPNLPKLVCHVHLKKSLSGSLPKVKNKSK 360
Qy 361 SPEEHLEEMKWMKSNKLTHTNPHIPKGGPPAKPKGKSHDKPLKAKGRSKGILNGQKSTGN 420
Db 361 SPEEHLEEMKWMKSNKLTHTNPHIPKGGPPAKPKGKSHDKPLKAKGRSKGILNGQKSTGN 420
Qy 421 SKSPKGLKTPKTKWKQMTLLDMAKGTQKMTAPRNSGGTPTRTSKPHKHLPPAALHLIA 480
Db 421 SKSPKGLKTPKTKWKQMTLLDMAKGTQKMTAPRNSGGTPTRTSKPHKHLPPAALHLIA 480
Qy 481 YYKENKREDKESALSVCISKTARLLSSDDRARLPEELRSLVOKYELLEHKKRWASKE 540
Db 481 YYKENKREDKESALSVCISKTARLLSSDDRARLPEELRSLVOKYELLEHKKRWASKE 540
Qy 541 EQRKEYLKKRBEELKKLKEKAKERREKEMLEERLEKQKRYEYDEQELTGKNLPAFRLVDTPE 600
```

```
Db 541 EQRKEYLKKREELKKLKEKAKEREKEMLERLEKQRYEDELTKGNLPAFLVDTPPE 600
Qy 601 GLPNTLFGDVAMVVEFLSCYSGLLLPDAQYPTITAVSLMEALSADKGGFLYLNRLVLI --- 657
Db 601 GLPNTLFGDVAMVVEFLSCYSGLLLPDAQYPTITAVSLMEALSADKGGFLYLNRLVILIQ 660
Qy 658 -LLQTLQDEIAEDYGELGWLKSEIPLTLHVSSELVRLCLRRSDVOEESGSDTDNKKDS 716
Db 661 TLLQTLQDEIAEDYGELGWLKSEIPLTLHVSSELVRLCLRRSDVOEESGSDTDNKKDS 720
Qy 717 AAPEDNEVDFLEKLETSFFELTSEEKLOILTALCHRLIMTYSVQDHMETRQMSAEL 776
Db 721 AAPEDNEVDFLEKLETSFFELTSEEKLOILTALCHRLIMTYSVQDHMETRQMSAEL 780
Qy 777 WKERLAVLKEENDKKAERKQKREMAKKNKENGKLVGKTKDRKKRIIVKFEPOVDTEAE 836
Db 781 WKERLAVLKEENDKKAERKQKREMAKKNKENGKLVGKTKDRKKRIIVKFEPOVDTEAE 840
Qy 837 DMISAVKSRLLLAIOAKEREIOEREMKVLERQAEERIRKHAAAEKAFQEGIAKAL 896
Db 841 DMISAVKSRLLLAIOAKEREIOEREMKVLERQAEERIRKHAAAEKAFQEGIAKAL 900
Qy 897 VMRTPIGTDNRNRYLWFSDEVPGLFIEKGWVHDSIDYRFNHHCKDHTVSGDEDYCPRS 956
Db 901 VMRTPIGTDNRNRYLWFSDEVPGLFIEKGWVHDSIDYRFNHHCKDHTVSGDEDYCPRS 960
Qy 957 KXANLKGNASMTQHGTAETAVAVETTPKQGNLWFLCDSQKELDELLNCLHPQGRSQ 1016
Db 961 KXANLKGNASMTQHGTAETAVAVETTPKQGNLWFLCDSQKELDELLNCLHPQGRSQ 1020
Qy 1017 LKERLEKRYQDIITHSLARKPNLGLKSCDGNQELLNFLRSDLIEVATRLQKGLGYBEE 1076
Db 1021 LKERLEKRYQDIITHSLARKPNLGLKSCDGNQELLNFLRSDLIEVATRLQKGLGYBEE 1080
Qy 1077 TSEFEARVLSLEKLDKDFGECVIALQASVTKKFLQGFMAQPKQKRLQSDSANTERVDSE 1136
Db 1081 TSEFEARVLSLEKLDKDFGECVIALQASVTKKFLQGFMAQPKQKRLQSDSANTERVDSE 1140
Qy 1137 KQWVEAKVASALEKWKTAIREAQTFSRMHVLGLMDACIKWDSAEANARCKVCPKKGED 1196
Db 1141 KQWVEAKVASALEKWKTAIREAQTFSRMHVLGLMDACIKWDSAEANARCKVCPKKGED 1200
Qy 1197 DKILILDECNKAFHLFCLRALPEYVDGEMQCPACOPATARRNSRGRNTYESASDSDSD 1256
Db 1201 DKILILDECNKAFHLFCLRALPEYVDGEMQCPACOPATARRNSRGRNTYESASDSDSD 1260
Qy 1257 DESDEEBEEEEEDYEVAGLRPRKTTIRGKHSVTPPAARSORRRPGKXPHSTRSQ 1316
Db 1261 DESDEEBEEEEEDYEVAGLRPRKTTIRGKHSVTPPAARSORRRPGKXPHSTRSQ 1320
Qy 1317 PKAPPVDDAEVDLVLQTKRSSRQSLQKCEIILHKIVKRFSPREPVRTDEADY 1376
Db 1321 PKAPPVDDAEVDLVLQTKRSSRQSLQKCEIILHKIVKRFSPREPVRTDEADY 1380
Qy 1377 YDVITHPMDFQTVQNKSCSGYSRVSQEFITDMQVFTNAEVNCRGSHVLSQVAKTEQCL 1436
Db 1381 YDVITHPMDFQTVQNKSCSGYSRVSQEFITDMQVFTNAEVNCRGSHVLSQVAKTEQCL 1440
Qy 1437 VLLHKLHPCHPVYRRKKFPDRLAEDGDSPEAVGQSRDEDRRSRAEIOEWLQDTS 1496
Db 1441 VLLHKLHPCHPVYRRKKFPDRLAEDGDSPEAVGQSRDEDRRSRAEIOEWLQDTS 1500
Qy 1497 LYSAKINSKDHNCFMMLVNTQFCMALTDVT 1527
Db 1501 LYSAKINSKDHNCFMMLVNTQFCMALTDVT 1531
```

RESULT 5
US-09-418-710-69
; Sequence 69, Application US/09418710
; Patent No. 6596482
; GENERAL INFORMATION:

```
; APPLICANT: Jones, Michael H.  
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR  
; FILE REFERENCE: 06501-042001  
; CURRENT APPLICATION NUMBER: US/09/418,710  
; CURRENT FILING DATE: 1999-10-15  
; PRIOR APPLICATION NUMBER: PCT/JP98/01783  
; PRIOR FILING DATE: 1998-04-17  
; PRIOR APPLICATION NUMBER: JP 9/310027  
; PRIOR FILING DATE: 1997-10-24  
; PRIOR APPLICATION NUMBER: JP 9/116570  
; PRIOR FILING DATE: 1997-04-18  
; NUMBER OF SEQ ID NOS: 73  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 69  
; LENGTH: 1525  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; US-09-418-710-69  
  
Query Match 99.0%; Score 7888; DB 4; Length 1525;  
Best Local Similarity 99.5%; Pred. No. 0;  
Matches 1521; Conservative 1; Mismatches 2; Indels 4; Gaps 4;  
  
Qy 1 MAPLLGRKPPPLVNPPLGPEPFTTIPHTQEAFTREYEARELYSERIWTCKTGSSQL 60  
Db 1 MAPLLGRKPPPLVNPPLGPEPFTTIPHTQEAFTREYEARELYSERIWTCKTGSSQL 60  
  
Qy 61 THKEAWEEOEVAELLKEEFPWYKLVLEMHVHNTASLEKLVDTAWLIMTKYAVGEEC 120  
Db 61 THKEAWEEOEVAELLKEEFPWYKLVLEMHVHNTASLEKLVDTAWLIMTKYAVGEEC 120  
  
Qy 121 DFEVGEKMLVKIKVIHPLEKVDSEATEKKSDGACDSSPKENSSQIAQHQKKTVV 180  
Db 121 DFEVGEKMLVKIKVIHPLEKVDSEATEKKSDGACDSSPKENSSQIAQHQKKTVV 180  
  
Qy 181 KEDEGRRESINDRARRSPKLPSTSLKGERKWAPPKFLPHKYDVKLQNEBKIIISNVPADS 240  
Db 181 KEDEGRRESINDRARRSPKLPSTSLKGERKWAPPKFLPHKYDVKLQNEBKIIISNVPADS 240  
  
Qy 241 LIRTERPPNKEIVRYFIRHNALRAGTGENAPMVVEDELVKYSLPSKFSDFLLDPYKYMT 300  
Db 241 LIRTERPPNKEIVRYFIRHNALRAGTGENAPMVVEDELVKYSLPSKFSDFLLDPYKYMT 300  
  
Qy 301 LNPSTKRNTGSPDRKPSKSTDNSSLSPLNPKLWCHVHLKKSLSGSPKVXNSKNSK 360  
Db 301 LNPSTKRNTGSPDRKPSKSTDNSSLSPLNPKLWCHVHLKKSLSGSPKVXNSKNSK 360  
  
Qy 361 SPEEHLEEMKQMSPNKLTNPHIPKGGPPAKPGKHSKPLKAKGRSGKILNGQKSTGN 420  
Db 361 SPEEHLEEMKQMSPNKLTNPHIPKGGPPAKPGKHSKPLKAKGRSGKILNGQKSTGN 420  
  
Qy 421 SKSPKKGLTKPTKMKQMTLLDMAGTKQWTRAPNSGGTPTSTSSKPHKHLPPAALHLTA 480  
Db 421 SKSPKKGLTKPTKMKQMTLLDMAGTKQWTRAPNSGGTPTSTSSKPHKHLPPAALHLTA 480  
  
Qy 481 YKKNKDRDKRSALSQVISTARTLLSSEDRARLPEELRSLVQRYELLEHKRWASMSB 540  
Db 481 YKKNKDRDKRSALSQVISTARTLLSSEDRARLPEELRSLVQRYELLEHKRWASMSB 540  
  
Qy 541 EQRKEYLKKRREELKKLKEKAKEREKEMLERLEKQRYEDELTKGNLPAFLVDTPPE 600  
Db 541 EQRKEYLKKRREELKKLKEKAKEREKEMLERLEKQRYEDELTKGNLPAFLVDTPPE 600  
  
Qy 601 GLPNTLFGDVAMVVEFLSCYSGLLLPDAQYPTITAVSLMEALSADKGGFLYLNRLVILIQ 660  
Db 601 GLPNTLFGDVAMVVEFLSCYSGLLLPDAQYPTITAVSLMEALSADKGGFLYLNRLVILIQ 660  
  
Qy 661 TLLQDEIAEDYGELGWLKSEIPLTLHVSSELVRLCLRRSDVOEESGSDTDNKKDSAAFE 720  
Db 660 TLLQDEIAEDYGELGWLKSEIPLTLHVSSELVRLCLRRSDVOEESGSDTDNKKDSAAFE 719  
  
Qy 721 DNEVQDFLEKLETSFFELTSEEKLOILTALCHRLIMTYSVQDHMETRQMSAELWKR 780  
Db 721 DNEVQDFLEKLETSFFELTSEEKLOILTALCHRLIMTYSVQDHMETRQMSAELWKR 780
```


Db 900 TP1GTDNRHNRVYLFSDSEVGLFIEKGMVHDSIDYRFNHHCKDHTVSGCEDYCPRSKKN 959
Qy 961 LGKNSMNTQHGATATEVAVETTPKQGNLWFLCDSQKELDELNCLNHPGIESOLKER 1020
Db 960 LGKNSMNTQHGATATEVAVETTPKQGNLWFLCDSQKELDELNCLNHPGIESOLKER 1019
Qy 1021 LEKRYQDI1HSIHLARKPNLGLKSCDGNQELLNFLRSLDLIEVATRLQKGLGVVEETSEF 1080
Db 1020 LEKRYQDI1HSIHLARKPNLGLKSCDGNQELLNFLRSLDLIEVATRLQKGLGVVEETSEF 1079
Qy 1081 EAEVISLEKLDKRGECVIALQASVIKKFLOGFMAPKQKRRKLOSEDSAKTEEVDEBKGMV 1140
Db 1080 EAEVISLEKLDKRGECVIALQASVIKKFLOGFMAPKQKRRKLOSEDSAKTEEVDEBKGMV 1139
Qy 1141 EAEVASALEKWTATREAOFTSRMEVLLGLMDACIKWMSAENARCKVCPKKGEDDKLI 1200
Db 1140 EAEVASALEKWTATREAOFTSRMEVLLGLMDACIKWMSAENARCKVCPKKGEDDKLI 1199
Qy 1201 LCDECNKAFLCLRLPALYEVDPGEWCQPACQAPATARRNSRGRNYTEESASEDESD 1260
Db 1200 LCDECNKAFLCLRLPALYEVDPGEWCQPACQAPATARRNSRGRNYTEESASEDESD 1258
Qy 1261 EEEEEEEEEEDYEVAGLRPRKTIIRGKHSVIPPAAARSGRRPKKPHSTRRSQPKAP 1320
Db 1259 EEEEEEEEEEDYEVAGLRPRKTIIRGKHSVIPPAAARSGRRPKKPHSTRRSQPKAP 1318
Qy 1321 PVDAAEDELVLQTKSSRRQSLQCEILHKKIVKRFSPFPVTRDRAEDYDVI 1380
Db 1319 PV-DAEDELVLQTKSSRRQSLQCEILHKKIVKRFSPFPVTRDRAEDYDVI 1377
Qy 1381 THPMDFTQVQKSCGSYRSVQEFELTDMKQVFNAAEVYVNCRGSHVLSVMVTEQCILVLL 1440
Db 1378 THPMDFTQVQKSCGSYRSVQEFELTDMKQVFNAAEVYVNCRGSHVLSVMVTEQCILVLL 1437
Qy 1441 HKHLPCHPYVRRKPKPPORLAEDGSEPEAVGQSRDEDRRSREAEIOEWLQDTSLY-S 1499
Db 1438 HKHLPCHPYVRRKPKPPORLAEDGSEPEAVGQSRDEDRRSREAEIOEWLQDTSLYAS 1497
Qy 1500 AKINSKDHNCFFMLVNTQFCMAITDTVT 1527
Db 1498 AKINSKDHNCFFMLVNTQFCMAITDTVT 1525

RESULT 7
US-09-949-016-7037
; Sequence 7037, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949, 016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7037
; LENGTH: 1540
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-7037

Query Match 96.3%; Score 7674; DB 4; Length 1540;
Best Local Similarity 99.3%; Pred. No. 0;
Matches 1474; Conservative 2; Mismatches 7; Indels 2; Gaps 1;

Qy 1 MAPLLGRKPPPLVNPPLGSEPPFTTIPHTQEAFTREYERBARLERYSERIWTCKTSSSQL 60
Db 58 MAPLLGRKPPPLVNPPLGSEPPFTTIPHTQEAFTREYERBARLERYSERIWTCKTSSSQL 117
Qy 61 THKEAWEEEOEVAELLKEEFPAPWYEKLVLEMHNTASLEKLVDTAWLEIMTKYAVGEEC 120
Db 118 THKEAWEEEOEVAELLKEEFPAPWYEKLVLEMHNTASLEKLVDTAWLEIMTKYAVGEEC 177
Qy 121 DFEVGEKMKLVKIVKHLEKVDDEATEKSGDAGCDSPSSDKENSSQIAQDHOKKETT 180
Db 178 DFEVGEKMKLVKIVKHLEKVDDEATEKSGDAGCDSPSSDKENSSQIAQDHOKKETT 237
Qy 181 KEDEGRRESINDRARRSPRKLPTSLKKGKRWAPPKFLPHKYDVKLQNEDKIISNPADS 240
Db 238 KEDEGRRESINDRARRSPRKLPTSLKKGKRWAPPKFLPHKYDVKLQNEDKIISNPADS 297
Qy 241 LIRTERPPNKEIVRPIRINALRAGTGENAPWVEDELVKCYSLPSKFSDFLLDPKYMT 300
Db 298 LIRTERPPNKEIVRPIRINALRAGTGENAPWVEDELVKCYSLPSKFSDFLLDPKYMT 357
Qy 301 LNPSTKRKNTGSDRKPSSKSTDNSSLSPLNPKLWCHVHLKSLSGSPLKVNKSNK 360
Db 358 LNPSTKRKNTGSDRKPSSKSTDNSSLSPLNPKLWCHVHLKSLSGSPLKVNKSNK 417
Qy 361 SPEEHLLEEMKMWSPNKLHFNH1PKGPPAKPKGKHSKPLKAKGRSGILNGQKSTGN 420
Db 418 SPEEHLLEEMKMWSPNKLHFNH1PKGPPAKPKGKHSKPLKAKGRSGILNGQKSTGN 477
Qy 421 SKSPKGLTKPTKMKQMTLLDMAKGTQKWTAPRNSGGTPTRTSSKPHKHLPPAALHLIA 480
Db 478 SKSPKGLTKPTKMKQMTLLDMAKGTQKWTAPRNSGGTPTRTSSKPHKHLPPAALHLIA 537
Qy 481 YKKNKORDEKESALSCVTSKTARLLSSDRARLPEELSLVQRYELLEHKKRWASME 540
Db 538 YKKNKORDEKESALSCVTSKTARLLSSDRARLPEELSLVQRYELLEHKKRWASME 597
Qy 541 EQRKEYLKKREBELKKLKEKAKERREKEMLESLVQRYELLEHKKRWASME 600
Db 598 EQRKEYLKKREBELKKLKEKAKERREKEMLESLVQRYELLEHKKRWASME 657
Qy 601 GLPNTLFGDVAMVVEFLSCYSGLLLPDAQYPTAVSLMEALSADKGGFYLNRVLVILQ 660
Db 658 GLPNTLFGDVAMVVEFLSCYSGLLLPDAQYPTAVSLMEALSADKGGFYLNRVLVILQ 717
Qy 661 TLQDEIAEDYGEGLMKLSEIPLTLHVSSELVRLCLRRSDVOBESGSDTDNDKDSAAPE 720
Db 718 TLQDEIAEDYGEGLMKLSEIPLTLHVSSELVRLCLRRSDVOBESGSDTDNDKDSAAPE 777
Qy 721 DNEVQDEFLKLETSSEKLIQILTALCHRLMTYSVQDHMETRQMSAELWKER 780
Db 778 DNEVQDEFLKLETSSEKLIQILTALCHRLMTYSVQDHMETRQMSAELWKER 837
Qy 781 LAVLKEENDKKAABKQKREKEMAKNKGKNGKNGKNGKNGKNGKNGKNGKNGKNGKNG 840
Db 838 LAVLKEENDKKAABKQKREKEMAKNKGKNGKNGKNGKNGKNGKNGKNGKNGKNGKNG 897
Qy 841 AVKSRLLAIQAKKEREIOEREMKYLKQAESEERIRKHKAABKQKNGKNGKNGKNGKNG 900
Db 898 AVKSRLLAIQAKKEREIOEREMKYLKQAESEERIRKHKAABKQKNGKNGKNGKNGKNG 957
Qy 901 TPIGTDNRHNRVYLFSDSEVGLFIEKGMVHDSIDYRFNHHCKDHTVSGDEDYCPRSKKN 960
Db 958 TPIGTDNRHNRVYLFSDSEVGLFIEKGMVHDSIDYRFNHHCKDHTVSGDEDYCPRSKKN 1017
Qy 961 LGKNSMNTQHGATATEVAVETTPKQGNLWFLCDSQKELDELNCLNHPGIESOLKER 1020
Db 1018 LGKNSMNTQHGATATEVAVETTPKQGNLWFLCDSQKELDELNCLNHPGIESOLKER 1077
Qy 1021 LEKRYQDI1HSIHLARKPNLGLKSCDGNQELLNFLRSLDLIEVATRLQKGLGVVEETSEF 1080
Db 1078 LEKRYQDI1HSIHLARKPNLGLKSCDGNQELLNFLRSLDLIEVATRLQKGLGVVEETSEF 1137
Qy 1081 EAEVISLEKLDKRGECVIALQASVIKKFLOGFMAPKQKRRKLOSEDSAKTEEVDEBKGMV 1140

Db	1138	EARVTSLEKLDKDFGECVIALQASVYKFLQGFAPKQKRLQSEDSAKTEEVDEBKXV	1197	
Qy	1141	BEAKVASALEKWKTAIREAQTSRMHVLLGMLDACIKWDMSAENARKVCPKKGDDKLI	1200	
Db	1198	BEAKVASALEKWKTAIREAQTSRMHVLLGMLDACIKWDMSAENARKVCRKKGDDKLI	1257	
Qy	1201	LCDECNKAHFLCLRALPYEVDGEGWQCPACOPATARRNSGRNRYTEESASDESDSD	1260	
Db	1258	LCDECNKAHFLCLRALPYEVDGEGWQCPACOPATARRNSGRNRYTEESASDESDSD	1317	
Qy	1261	EEEEEEEEEEEDYEAGLRPRKTIIRGKHSVPPAARSGRRPGKPHSTRSQPKAP	1320	
Db	1318	EEEEEEEEEEEDYEAGLRPRKTIIRGKHSVPPAARSGRRPGKPHSTRSQPKAP	1377	
Qy	1321	PVDDAEVDELVLQTKRSSRRQSLQKCEIILHKIVKRYFSPWPPREPVTRDAEDYDVI	1380	
Db	1378	PVDDAEVDELVLQTKRSSRRQSLQKCEIILHKIVKRYFSPWPPREPVTRDAEDYDVI	1437	
Qy	1381	THPMDPQTVONKSCSGSYRSVQEFPLTDMKQVFTNAEVYNCRSHVLSCHVKTQCLVILL	1440	
Db	1438	THPMDPQTVONKSCSGSYRSVQEFPLTDMKQVFTNAEVYNCRSHVLSCHVKTQCLVALL	1497	
Qy	1441	HKHLPGHPVRRKRRKFFPRLAEDEGDSPEAVGOSRDEDRSRE	1485	
Db	1498	HKHLPGHPVRRKRRKFFPRLAEDEGDSPEAVGOSR--GRRQKK	1540	
RESULT 8				
US-09-418-710-1				
; Sequence 1, Application US/09418710				
; Patent No. 6596482				
; GENERAL INFORMATION:				
; APPLICANT: Jones, Michael H.				
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR				
; FILE REFERENCE: 06501-042001				
; CURRENT APPLICATION NUMBER: US/09/418,710				
; CURRENT FILING DATE: 1999-10-15				
; PRIOR APPLICATION NUMBER: PCT/JP98/01783				
; PRIOR FILING DATE: 1998-04-17				
; PRIOR APPLICATION NUMBER: JP 9/310027				
; PRIOR FILING DATE: 1997-10-24				
; PRIOR APPLICATION NUMBER: JP 9/116570				
; PRIOR FILING DATE: 1997-04-18				
; NUMBER OF SEQ ID NOS: 73				
; SOFTWARE: FastSeq for Windows Version 4.0				
; SEQ ID NO 1				
; LENGTH: 1674				
; TYPE: PRT				
; ORGANISM: Homo sapiens				
US-09-418-710-1				
Query Match 11.7%; Score 929.5; DB 4; Length 1674;				
Best Local Similarity 23.1%; Pred. No. 8.1e-58;				
Matches 395; Conservative 264; Mismatches 529; Indels 525; Gaps 67;				
Qy	3	PILGRKPP-----PLVNPLGCEBPFPTIPTHQAFRTREYBEARLERYSTRCKSTGSS	58	
Db	152	PLLHRKPPVQRKPPADLRDPDEEVFY-CVKTNBIFRHYDDFFERTILCNLSVMSCAVTGRP	210	
Qy	59	QLTHKEAWEQEVAELLKEEPPAWVEKLVLEMVH---HNTASLEKLVDTAWLEIMTKY	114	
Db	211	GUTYQALESEKKARQNL-QSFP---EPLIIPVLYLTSITHRSLRHEICDDIFAYVKORY	266	
Qy	115	AVGEECDDFVGKEMKLVKIVKHPLEKVDDEATEKKSDGACDSPSDKENSQIAQDHQ	174	
Db	267	FVEETVEVIRNNGARLQCTILEVLP-----PS---HQNGFANGHV	303	
Qy	175	KK---ETVKEDEGRRESINDARRSPRLPTSLKKGGRKWAAPPKPLPHKYDY----	KLQ 227	
Db	304	NSVDGETIISDSDSETQS-----CSFQNGKKKDAIDPLL-FKYKVQPTKKBEL	351	
Qy	228	NEDKIISNPADSLINTERPPNKEIVRYFIRHNALRAGTGENAPWVEDELVKYKSLPS-	286	

Db	352	HESAI---VKATQISRRKHLFSRDKLKLFLKQHC-----EPQEGVIK---IKASSLSY	399	
Qy	287	-----KPSDFLLDPYKYMTLAPSTIRKNTGSPDRKPSK---SKTDSLSLSPFLPKLWC	338	
Db	400	KIAEQDFSYFFPDDPPTFIFSPANRRG-----RPPKRIHISOEDN-----	440	
Qy	339	HVHLKKSLSGSPPLKVNSKNSKSPDEHLEEMMKMSPNKLHTNPHI PKKGPAPKPKGKS	398	
Db	441	-VANKQTIA-----SYRSKATKER-----	458	
Qy	399	DKPLKAGKSGIILNGQSTGNSKSPKGLKTPKTKMQMTLLDMAKGTQKMTAPRNSG	458	
Db	459	DKLLK-----QEEMSLAFKAKLKR-----	479	
Qy	459	GTPTSTSPKHKLPPAALHLIAYYKENDKRSALSCVISTARTLLSSDDEARLPEBL	518	
Db	480	-----EKADALEAKKKEDEKCKR-----	502	
Qy	519	RLVQKRYELLEHKRWASMSSEBQKRYLKKRBEELKKLKEKAKEREKEMLERLEKOK	578	
Db	503	KKIVEE-----RLKKBEKRLKVEREKERKLE--EKRTVEYLVKOWSKPR	549	
Qy	579	RYEDQELTG-KNLPAFRLVDTPEGLPNTLFGDVAMVVEFLSCYSGILLPDAQY--ITAV	635	
Db	550	--EDMECDLDELPEPTPVKT--RLPPEIFGDALMVLEFLNAFAGELFDLODFPDGVTL	605	
Qy	636	SLMEAL-SADKGG-----FLYNRVLVILQTLQDEIAE-----DYBELGMLSEIPLT	684	
Db	606	VLEEALVNGDSGLCELLFFLTAFOAIAEE--EEEVAKEQLTADTKGCSLSLDLD	663	
Qy	685	LHSVSELVRLCLRRSDVQESSEGSTD-----DNKDSAAFEDNEVDQEFLEKLETSF	737	
Db	664	SCTLSEILRLHLASGADVTSANAKRYOKRGGFATDDACMELSLNPSLVKKLSSTSV	723	
Qy	738	FELTSEEKLOILTALCHRLMTYSVQD-----HMETRQMSA	774	
Db	724	YDLTPEGKMKILHALCGKLLTLVSTRFDIEDYVDILROAKQEFFRELKABQHRKEREAAA	783	
Qy	775	ELWKERLAVLKEENDKKAERKQKXEMEAKN-----KENGKVENGLGKTDKKRIV	825	
Db	784	RIRKKEEKLKEQEKMEKQEKLEKDEQRNSTADISIGEEEREDFTSIESKDETK--	841	
Qy	826	KFEPQVDTAEADMISAVKSRLLAIQAK-----KEREIQE-----REMKVLAQEA-E	872	
Db	842	ELDQDMFTDEDDPGSHKGR-----RGRKGNGKFEFTRQEQINCVTRRELLTADBEELK	897	
Qy	873	EERIRKHAARAFQEGIAKAKLVMRRTPIGTDRNHNRYWLFSDVDFGLFIEK---GWV	929	
Db	898	QEHQKEKELLEK-IQSAIACNIF-----PLGRDRMYRRYWF-PSIPGLFIEEDYSGLT	951	
Qy	930	HDSI-----DYRFNHCKDHTVSGDEDCYCPRSKKN-LGKNASMTQHTGATEVAVETTT	983	
Db	952	EDMLLRPPSSFONNVQSQDPQVS-----TKTGEPLMSESTSIDQG-PRDHSVOLPK	1002	
Qy	984	PQOGNLFCDQSOKELDELLNCLHPQGIRESQLAKERL--EK-----RYODIHSFH	1033	
Db	1003	PVHKPNRCFYSSCQOLDOLIEALNSGRHSALAKETLLQEKSRICAQLARSE--EKPH	1060	
Qy	1034	LARKENLGLK-----SCDGNQ-----ELNLFRLSDLJEVATRLQKGGIGVYEET-	1077	
Db	1061	FSDKQPPSKPYSGRSSNAVDPQMAEKQLELRDLFLDIEDRIYQGTGAIKVTD	1120	
Qy	1078	-----SEFEA---RVISLEKL-----KDFGECVIALQASVTKKFLQGMAPKQKR	1120	
Db	1121	RHWRSALESGRYELLSBENKENGIIKTVNEDVEEMEIDEQTKVIVK--DRLLGKKTETP	1178	
Qy	1121	KLQSDSAKTEVDE-----EKKWVEEAKVAS-----ALEKWKTAI	1156	
Db	1179	STVTNASTPQSVSSVHVYLAVALFQIEQGIERRFLKAPLADSDSGRSYKTVLDNRRESL	1238	
Qy	1157	REAOFTSRMHVLLGLMDACIKWDMSAENARCKVCPKKGDEDDKLILCDECNKAHFLCLRP	1216	

Db 1239 LSSASLSQVFLHLSTLDRSVINSGKILNARCKICRKKGDAENMVLCDGCDRGHTTYCVRP 1298
Qy 1217 ALYVEPDGWCQACOPATARRNSRGRNTYESASEDSEDESDEEE-----EEEBEBEE 1272
Db 1299 KLKTVPEGDFWCEPCRPKQCRRLSPRQPSLESDEDEVEDSMGDEDEVDGDEEQSGSEE 1358
Qy 1273 EDEYV-----AGLRAPRXTI-----RKHSHVPPAARSRRGPKKPHSTRR 1314
Db 1359 EYEVEQDEDDQOEBEVSLPKRGRPOVRLPVKTRGKLSSTSSRQGOQEPGRYPSSQO 1418
Qy 1315 SOPK-----APPVD-----DAEYDEL-----1330
Db 1419 STPKTTVSSKTSGLRKINSAPPTETKSLRIASRSTRSHGHPLOADVVELLSPRKRRG 1478
Qy 1331 -----VLQTKRSSRQSL-----1343
Db 1479 RKSANNTSPNSPNFNRVIATKSSQSRSVNIASKLSLQESKRCRKRQSPSPVPT 1538
Qy 1344 -----ELQKCEILHKIVKRFSPFPREPVTTRDEADYDVITHPMDPTQVQ 1390
Db 1539 LGRSSRGOGVHLSAQFLVVELVRHDDSWPFLKLSKIQVPDYDIKKPIALNIIR 1598
Qy 1391 NKCSGSGYSVOEFLTDMKQVFTNAEYVNCGRS 1423
Db 1599 EKNKCEYKLASEFIDDIELMFNSCFEYINPRNT 1631
RESULT 9
US-09-839-479-1
; Sequence 1, Application US/09839479
; Patent No. 672722
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042002
; CURRENT APPLICATION NUMBER: US/09/839,479
; CURRENT FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1674
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-839-479-1
Query Match 11.7%; Score 929.5; DB 4; Length 1674;
Best Local Similarity 23.1%; Pred. No. 8.1e-58;
Matches 395; Conservative 264; Mismatches 529; Indels 525; Gaps 67;
Qy 3 PLLGRKPF---PLVNPLGEPFFTIPTQAFRTREVEARLERYSERIWTCKSTGSS 58
Db 152 PLLHRAPFVRQPPADLRDEEVFY-CKVTNELFRHYDDFFERTILCNLSLWSCAVTGRP 210
Qy 59 QUTHKEAWEEQEVALLKEEFPWYKELVLEMVH-----HNTASLEKLVDTAWLEIMTKY 114
Db 211 GLTYQEALSEKKARQNL-QSFP---EPLIIPVLYLTSLTHRSRLHEICDDIFAYVKORY 266
Qy 115 AVGECDCFEVGKMKLVKVIKHPLEKVDDEATEKKSDGACDPSPSDKENSQIAQHQ 174
Db 267 FVEETVEVRNNGARLQCTILEVLP-----PS-----HQNGFANGHV 303
Qy 175 KK---ETVYKEDEGRESINDRARRSPRLPTSLKKGKRWAPPKFLPHKYDV-----KLQ 227
Db 304 NSVDGTIIISDDSETQS-----CSFQNGKKDAIDPLL-FKYVQVPTKREL 351

Qy 228 NEDKIISNPADSLIRTERPPNKEIVRYEIRNNAALRAGTGENAPWVVEDELVKKYSLPS- 286
Db 352 HESAI---VKATQISRRKHLFSRDKLFLKQHC-----EPOEGVIK---IKASSLSTY 399
Qy 287 -----KFSDFLLDPYKMTLNPSTKRNKNGSPDRKPSK---SKTDSNSSLSSPLNPKLWC 338
Db 400 KIAQDFSYFFPDPTFFISPANRRG-----RPPKRIHISOEN-----440
Qy 339 VHYLKKLSGSPKLVKNKSNKSPBEHLEEMKMTMSPNKLHTNFHFKKGPAPKPGKHS 398
Db 441 -VANKQTLA-----SYRSKATKR-----458
Qy 399 DPLPAKAGRSKILNGQKSTGNSKPKGLKTPKTKYKMKQMTLLDMAKGTKQKTRAPNSG 458
Db 459 DKLLK-----QBEKSLAFKAKLKR-----479
Qy 459 GTPRTSSKPHKLPPAALHLIAVYKENKDRDKRSALSVCISKARLLSSDSEARLPEEL 518
Db 480 -----EKADALEAKKKEKEDKCKR-----EEL 502
Qy 519 RSLVQKRYELLEHKRWRWASBEQRKYLLKKREELKKLKEKAKERREKEMLERLEKQK 578
Db 503 KKIVEEE-----RLKKKEKERLKVEREKEREKLE---EKKRYVEYLKQSKPR 549
Qy 579 RYEDQELTG-KNLPAPRLVDTPEGLPNTLFGDVAMVVFSLCYSYGLLLPDAQYP--ITAV 635
Db 550 --EDMECDLDELPEPTPVKT--RLPPEIFGDALMVLEFLNAFELFDLQDFPDGVTLE 605
Qy 636 SLMEAL-SADKGG-----FLYLNRLVILLQTLQDEIAE---DYGELGMKLSIPT 684
Db 606 VLEEALVGNDSGPICELLFFFLTAIFOAIEE---EEEVAKGQLTDADTKGCSLKLDDLD 663
Qy 685 LHSVSELVRLCLRRSDVQBESESGSDTD-----DNKDSAAFEDNEVDQEFLEKLTSEF 737
Db 664 SCTLSEILRLHLASGADVTSANAKYRYOKRGGFATDDACHELRLSNLSLVKLSSTSV 723
Qy 738 FELTSEEKQILTALCHRLMTYSVOD-----HMETROOMSA 774
Db 724 YDLTPGKMKILHALCGKLLTLVSTRDFTEDYVILRQAKQBFRELKAEQHRKEREEAAA 783
Qy 775 ELWKERLAVLKEENDKRAEKOKRMEAKN-----KENGKVENGLGKTDRKKRV 825
Db 784 RIRKREKLEKQEQOKMEKQEKLEDEQORNSTADISIGEEREDFTSIESKDTQK-- 841
Qy 826 KPEPOVDTEAEDMISAVKSRRLLAIOAK-----KERETQE-----RBMVKLERQA-E 872
Db 842 ELDDQMFTEDEDDPGSHKGR---RGKEGQNGKFEFTQEQINCVCYTRELLTADEEALK 897
Qy 873 EERIRKHAABAKAFOEGIAKAKLVWRTPITGTDNNHNYWLFSDSEVPGLFIEK---GWV 929
Db 898 QEHQKEKELEK-IQSAIACNTNIF---PLGRDRMYRRYWF-PSIPGLFIEEDYSGLT 951
Qy 930 HDSI-----DYRFNHCKDHTVSGDEDCYCPRSKAN-LGKNASMTQHGCTATEVAVETIT 983
Db 952 EMLLPFRSSFNQVQSDQPVQS-----TKTQELMSESTSNIDQG-PRDHSVOLPK 1002
Qy 984 PRQGNLWFLCDSQKELDELNCLHPQGIRESQKLERL--EK-----RYODIHSIH 1033
Db 1003 PVHKPNRWCFYSSCEQLDOLIEALNSRGHRESALKETLLOEKSRICAQLARSE--EKPH 1060
Qy 1034 LARKPNLGLK-----SCDGNQ-----ELNPLRSDLIEVATRLQGGYGYEET- 1077
Db 1061 FSDKFPQDSKPTYSRGRSSNAYDPSQMAEKQLELRDLRDLIEDRIYQGTILGAIKVTD 1120
Qy 1078 -----SEFEA---RVISLEKL-----KDFGECVIALQASVIKKFLOGPWAPKQKR 1120
Db 1121 RHIVSALESGRYELLSEENKENGIIKTVNEDVEEMEIDEQTKVIK--DRLLGIKTETP 1178
Qy 1121 KLQSDSASKEVDE-----EKKWVEEAKVAS-----ALEKWKTAI 1156
Db 1179 STVSTNASTPQSVSSVHYLMALFOIEQIERRFLKAPLDASDSGRSYKTVLDRWRESL 1238
Qy 1157 REAQTFSRMHVLLGLMDACIKWDMSAENARCKVCPKKBDDDKLILCDECNKAFHFLCLR 1216

Db	1236	SLUSSASLQVFLHLSLDRSVIWSKILNARCKICKKGDAENMWLDCDGRGHITYCV	12395
Qy	1215	RPALYVPGDEWQCPACQATARNRGRNYTESASESDEDESEDEEE- - - - EEEEE	1270
Db	1296	RPKLKIVPEGDWFCPECRPQRCRLRSFRQPSLESDSEDEVEDSMGGEDDEVDGEERGQS	1355
Qy	1271	BEEDYEV- - - - - AGLRLRPKTI- - - - - RGKHSVTPPAARSRRRPGKKPHST	1312
Db	1356	EEBEYEVEQDEDDSQBEIEVSLPKRGPRQVRLPVKTKGKLSFSFGQQOEGRYPSSR	1415
Qy	1313	RRSOPK- - - - - APPVD- - - - - DASVDL- - - - -	1330
Db	1416	QOSTPKTIVSSKGRSLRKINSAPPIETKSLIASRSTRSHGHPLOADVVELLSPRKR	1475
Qy	1331	- - - - - VLOTGRSRRQSL- - - - -	1343
Db	1476	RGRKSANNTPENSFPNFPNFRVIATKSSQEQSRSVNIASKLSLOESGKRRCKRQSPESP	1535
Qy	1344	- - - - - ELQKEEILHKIVKVFWSFPMPPEYVTRDEADYDYVITHPMDFQT	1388
Db	1536	VTLGRSSGROGGVHELASAEOQLVVELVRHDDSWPFLKLVSKTQVPDYDIKKPTALNI	1595
Qy	1389	VQNKSCGSYRSVQEFLLDMKQVFTNAEYVNCGRS	1423
Db	1596	IREKVNCKEYKLASEFIDDIELMFSCFNCPNPNPT	1630

```

RESULT 12
US-09-513-999C-7706
; Sequence 7706, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert, A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.

```

Query Match	8.0%;	Score 637;	DB 4;	Length 128;	
Best Local Similarity	99.2%;	Pred. No. 2.4e-38;			
Matches 127;	Conservative 0;	Mismatches 1;	Indels 0;	Gaps 0;	
QY	676	MKLSEIPTLHLSVSELVRLCLRRSDVQEESEGS	DTDDNKDSAAFDNEVQDFLEKLETS	735	
Db	1	MKLSEIPTLHLSVSELVRLCLRRSDVQEESEGS	DTDDNKDSAAFDNEVQDFLEKLETS	60	
QY	736	EFPELTSEKLOITLALCHRLIMTYSVQDHMETFO	QOMSALWKERLAVLKEENDKKRAEK	795	
Db	61	EFPELTSEKLOITLALCHRLIMTYSVQDHMETR	QQMSALWKERLAVLKEENDKKRAEK	120	
QY	796	QXKREMA 803			
Db	121	QXKREMA 128			
RESULT	13				

```

US-09-418-710-21
; Sequence 21, Application US/09418710
; Patent No. 6596482
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042001
; CURRENT APPLICATION NUMBER: US/09/418, 710
; CURRENT FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 21
; LENGTH: 1972
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-418-710-21

```

Query Match	6.7%	Score 532;	DB 4;	Length 1972;
Best Local Similarity	18.1%;	Prod. No. 4.2e-29;		
Matches	354;	Conservative 256;	Mismatches 599;	Indels 750; Gaps 68;
Qy	8	KPFLVNPLPGEPFF-FTIPH----	-----TQEAFFRTREEYEALERYSERIWTCKST	55
Db	189	KPLSLVNOAKETYMKLVIPSPDVLKAGNKNTSESSLLTSELRSKEQY----	-----KQA	241
Qy	56	GSQQLTHKEAWEEQEAELLK-----	EEFPWYKLVLEMVHN	97
Db	242	FPQQLKQESSLKVKVIAALSNPKATSSPA-HPKQTLLENHPNPFLLNALLGNHOPNG	300	
Qy	98	SLKELVDTAWLMTVKVANGCEDPEV-----	-----	124
Db	301	VIOGVIQEAFLATLTKTKQSKINENIAAASPTFPSPVNLSTSGRRTPNGQTPVMPSPAS	360	
Qy	125	-----GKEKWL-----KVKIVKI-----HPL-----	-----	146
Db	361	PILHSQGEKAVANNVNPVKTHSHSPAKSLVQRFGTSDIPSSKXSDSEDSNEDEEDDE	420	
Qy	147	ATBKSDGACDPSSEKENSQIAQHOKKETVVKEDGRRESINDPARRSPKLPITSLK	206	
Db	421	EEDEEDDEDSDQSESDSNSSESTEGSEEDDDDKQOQESDSD-----	-----T	467
Qy	207	KGRKWAAPPKPLPHKYDVKLONDKIISNV--PADSLIRTERPENKEIVRYFIHNAALRA	264	
Db	468	EGE-----KYSMKL-----NKTSSVKSPMSLTGHSTPRNLHIAK-----	-----	503
Qy	265	GTGENAPWVVEDBLVKYSLPSKESDFLDLPYKYMTLNPSTKRNKNTGSPDRKPSKSK-T	323	
Db	504	-----APGSAPAALCESQSPA-----	-----FLGTSSSTL-----TSSPHSGTSKRRVVT	544
Qy	324	DNSLSPLNPKLWCHVHLKKSLSGSLPKVKNKSNKSP-----	EEHLEEMKMKMSPN-----	376
Db	545	DERELRIPLEYG-WQRETRIRNFGG--RLQGEVAVYAPCGKKLQRPVEVIKYLRSNGIM	600	
Qy	377	-----KLHT-NFHIPIPKGP-----	-----PAKKPG	395
Db	601	DISRDNFSFAKIRVGDIFYARDGPQEMQWCLLKEDVPIPRAMEGRRGRPNPDRQRA	660	
Qy	396	KHSDKPLKAKGRSKI-----	-----LNQKSTGNS	421
Db	661	REESRMRRRGRPNPVNGNAEFLDNADAKLRKLQAEIARQAAQIKLLRKLQEQEARVA	720	
Qy	422	KSPKK--GLKTPTKQKQMTLLDMAGTKQKWRAPRNSGGTPTRTSSKPHKHLPPAALHLI	479	
Db	721	KEAKQQAIIWAEEKQKEQIKIMQOQEKIKRIQO-----	-----IEMKEL--RAQOIL	769
Qy	480	AYYKENKDRBKRSALSCVSIKTLARLSSSEDRARLPEELRSLVOKRYELLEHKRWASMS	539	

Db 770 EAKKKKEE-----AANAALLEAEKRIK-EKEMR-----ROQAVLLKHQER----- 809
Qy 540 EQRKEYLKKRBEELKKLKEKAKERREKEMLEKOKRYEDQEL---TGKIL----- 590
Db 810 ERRROHMLMKAMEARKAEKLEKQKDEKRLNKKERLEQORRLELEMAKELKKPNE 869
Qy 591 -----PAPRLVDTPB-GLPNTLFGDVAMVVEFLSCYSGLLLPDQOYPITAVSLMEALS 642
Db 870 MCLADQKPLPELPRIFCLVLSGTFSDCLMVQVFLRNFGLVGFVDVNIIDVPLNSVLQEG 929
Qy 643 ADKGGFL-YLNRVILLQTLLODE-IAEDY-----GELGMKLSRPLTLHSVSELVRLCLR 697
Db 930 LNIQDSNGEVQDLVRLLSNAVCDPGLITGYKAKTALGEHLNVGNRDNVSEILQIFW- 988
Qy 698 RSDVQBESEGSDDDNKNSAAFEDNEVDQEFLEKLETSSEFFELTSBEKLQILITALCHRI 757
Db 989 -----EAHCGQTELTSKATKAFQAHTPAQKASVLAFLINELA 1026
Qy 758 MYSVQDHMETROQSAELWELAVLKEENDKKRAEKQKREKEMAKNKENGKVENG--- 814
Db 1027 CSKSVVSEIDKNIDYMSNLRDKWV-----EGKRLRLRIIHAKKTKGRDTSGGIDILGERQ 1082
Qy 815 --LG-----KTRKKRIIVKFEQVDEAEDMISAVKSRLLALIAOAKKEREIOE--REMK 864
Db 1083 HPLGTTPTGKRRKRGDSDYDDDDSDSD-----QGEDEDEEDKEDQK 1129
Qy 865 VKLQAEAEERIRKHAKEA-----AFQEGIAKAKLVMRRTPTGDRNHR 911
Db 1130 GKXTDICEDEGDOAASVEELEKQIEKLSKOOSQYRKLFDAHSLSRSMVFGPDYRRR 1189
Qy 912 YMLFSDVEPGLFIE-----KGWVH-----DSIDYRFNHC 941
Db 1190 YMIL-PRCGIFVEGMESGEGLEAEIAKEREKLLKABSVQIKEMFETSGDSLNCSTNDC 1248
Qy 942 -----KDHT---VSGDEDYCRSK-----KANLGNA----- 965
Db 1249 EOKEDLKEKDNMLFQKPGSFKLSJLEVAKMPPESEVMTPKNAGANGCTLSYQNSG 1308
Qy 966 --SMTQHTGATEVAVE-----TTPPKQQLNWL- 993
Db 1309 KHSLSGVQSTATQSNVEKADSNLNTGSSGPKFYSLPNDQLLTLEKRNQ--WFSL 1366
Qy 994 --CD----- 995
Db 1367 LPRTPCDDTSLTHADMSTASLVTPOQPPSKSPSPAPLGGSAQNPVGLNPFALLSPLQV 1426
Qy 996 ----- 995
Db 1427 KGVSMMLQFCGWPTGVVTSNIPFTLSVPSLGLGSEGNGNSFLTNSVASKSESPV 1486
Qy 996 -----SOKELDELNCLHPOGIRE 1015
Db 1487 PONEKATSAQPAAEVAKVPDPSPKPIPEEMQFGWRIIDPEDLKALLKVLHLGIREK 1546
Qy 1016 QKLERLEKRYQDIISHILARKPNLGLSKSCDGNQELLNFLRDLIE---VATRLQKGGIG 1072
Db 1547 ALQKQIQK-HLDYITQACLKNK-DVAIIELNENE--NQVTRDIVENWSVEQAMEMDLS 1602
Qy 1073 VYEETSEPEARVISLEKDFECVITALQASVIKKFLOFMAPQO----- 1117
Db 1603 VLOQVEDLERRVAS-----ASLO-----VKGMCPPEASEREDLVYFHKSF 1644
Qy 1118 -----KRRKLQSEDSAKTEEVDEEKKWVE--EAKVASALEKKW 1153
Db 1645 TKLKEHDEGFTGEDESSAHALERKSDNFDLATVTLADLERNIERRIEEDTAPGLRWVR 1704
Qy 1154 TAIREAQTPSRMHVLGLMLDACIKWDMSAENARCKVCPKKGDDDKLILDECNKAFHFLC 1213
Db 1705 RALSEARSAQVALCIQLOKSLAWKSIWKVYQICRKGDNELLILLDCGDKGCHTYC 1764
Qy 1214 LRPALYEVDPGWQCPACOPATA-----RNSRGRNVTESASE-----DSEDESD 1260
Db 1765 HRPKITTIIDGDMFCPACTAKASGOTLKI KLVKVGKKTNESKGGKVTLTGDEDEDS- 1823

Qy 1261 EEEEEEEEEEDYEAGLRRLRPRTIRGKHSHVPPAARSGRRPGKPKHSTRRSQKAP 1320
Db 1824 -----ASTSSSLKRGKNKDLQKKEMEENTS 1847
Qy 1321 PVDVDAEVDLVLOTKSSRRQSLBQCEIHLHKIVKRYFSWPPRPVTRDEAEDYDVI 1380
Db 1848 --INLSKQSFSTSVKPKRDDSKDLALCSMILTEMETHEDAWPFLLPVNLKLVPGYKVI 1905
Qy 1381 THPMDPOTVONKSCSGSYRSVOEFLTDMKQVFTNAEVYN 1419
Db 1906 KXPMDFSTIREKLSGQYENLETFAVDLVRVDFDCETFN 1944
RESULT 14
US-09-839-479-21
; Sequence 21, Application US/09839479
; Patent No. 6727222
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042002
; CURRENT APPLICATION NUMBER: US/09/839,479
; CURRENT FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 21
; LENGTH: 1972
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-839-479-21
Query Match 6.7%; Score 532; DB 4; Length 1972;
Best Local Similarity 18.1%; Pred. No. 4.2e-29;
Matches 354; Conservative 256; Mismatches 599; Indels 750; Gaps 68;
Qy 8 KPFLVNLPLGPEPF-FTIPH-----TOEAFRTREYEYEARLEYSERIMTCKST 55
Db 189 KPLSLVNAQAKTYMKLIVPSPDVLKAGNVTSESSLLTSELRSKREY-----KQA 241
Qy 56 GSSQLTHKEAWEEEOEVAELK-----BEFPAYEKLVLVMVHN-----TA 97
Db 242 PPSQLKKQESSKSLKVVIAALSNPKATSSSPA-HPKQTLNHNHPFLTNALLGNHPNG 300
Qy 98 SLEKLVDTAWLAIMTYAVGEECDFEV----- 124
Db 301 VTQSVIQEAPLALTATTKMQSKINENIAAASPTFSSPVNLTSGRRTPGNOTVMPSPAS 360
Qy 125 -----GKEKML--KVKIVKI---HPL-----EKVDEE 146
Db 361 PTLHQQGKEKAVNNVNPVKQHHHPAKSLVEQPRGTDSIPSSKQSDSEDEEEDDE 420
Qy 147 ATEKKSQDCAQSPSSDKENSSQIAQDHQKKTVMKEDEGRRESINDRARRSPKRLTSLK 206
Db 421 BEDEEDEDDESDSQSESDSNSSEDTGSEEEEDDDKQDSDSD-----T 467
Qy 207 KGERKWAAPPKFLPHKYDVKLQNEDKIISNV---PADSLRTERPPNKEIVRYFIRHNALRA 264
Db 468 EGE-----KTSMKL---NKTTSVKSPMSLTCGHSSTPRNLHIAK----- 503
Qy 265 GTGENAPWVVEDELVKYSLPSKFSDFLLDPKYKNTLPSTKRYKVTGSDPKPSKSK-T 323
Db 504 -----APGSAAPALCESQSPA-----FLGTSSSTL---TSSPHSGTSKRRVT 544
Qy 324 DNSSLSSPLNPKLWCHVHLKKSLSGSLKSGPLKVKNSKNSKSP-----EBHLEEMKMWSPN--- 376

Db	545	DERELRIPLEYG-WQRETRIRNFGG---RLQGEVAYYAFCGKKLRQYPEVIVKYLRSNGIM	600	Db	1547	ALQKQIQK-HLDYITQACLKNK-DVAIIENENEE--NOVTRDIYENWNSVEEQAMEMDLS	1602
Qy	377	-----KLHT-NFHIPKKGPP-----PAKRP	395	Qy	1073	YVEETSEFEARVISLEKLDKDFCECVIALQASVKKFLOGEMAPKQ-----	1117
Db	601	DISRDNFSFAKIRVGDGYEARDGPOEMQWCLLKEEDVIPRIRAMEGRRRPDPNDQRA	660	Db	1603	VLQOVEDLERRVAS-----ASLQ-----VKGWMCPEPASEREDLVYFEHKS	1644
Qy	396	KHSDKPLKAGRSKGI-----LNGOKSTGNS	421	Qy	1118	-----KRRKLOSEDSAKTEEVEDEEKQWVE--EAKVASALEKWK	1153
Db	661	REESRMRKGRPPNVGNAEFLDNADAKLLRLQAEIARQAQIKLRLKQKQOARVA	720	Db	1645	TKLCKEHGDEFTGEDESSAHALERKSDNPLDIAVTRLADLERNIERRIEEDTAPGLRVWR	1704
Qy	422	KSPKK--GLUKTPKTKOMQMTLLDMAGTKQMTTRAPRNSGGTPTSTSSKPHKHLPPAALHLI	479	Qy	1154	TAIRAEQFTSRMHVLLGMLDACIKWDMASAEANARCKVCPKKGDEDDKLIILCDECNKAFLPC	1213
Db	721	KEAKQQAIAAEKKEKQEQIKIMKQEQEKIRIQO-----IRMEKEL--RAQQLL	769	Db	1705	RALSARGAAQVALCIIQOLQKSIAMEKSTIMKVYCYICRKGDNNEELLLCDGCDKGCHTYC	1764
Qy	480	AYYKENKREDKRSALSCVISTARLLSSEDRARLPEELRSVLQKRYELLEHHKKWASMS	539	Qy	1214	LRPALYEPVDGQWQCPACQAPATA-----RNSRGRNYTESASE-----DSEDDSD	1260
Db	770	EAKKKKEE-----AANAKLLEAEKRIK-EKEMR---RQOAVLLKHQER-----	809	Db	1765	HRPKITITPDGDFWFCPACIAKASGQTKIKLHVKGKKTNESKKGKKVTLTGDTDEDS-	1823
Qy	540	EQORKEYLKKREELKKLKEKAKERREKEMLERLEKQKRYEDQEL---TGKNL-----	590	Qy	1261	EEEEEEEEEEEDYEVAGLRRLRPRKTIRGKHSVIPPAARSGRRPGKKPHSTRSQPKAP	1320
Db	810	ERRROHMLMKAEARKKAEERUKQEKRDCKRLNKERKIQORRLEMAKELKPNED	869	Db	1824	-----ASTSSSLKGNKDLQKRKMEINTS	1847
Qy	591	-----PAFLRVDTPE-GLPNTLFGDVAMVVEFLSCYSGLLPDAQYPIITAVSLMEALS	642	Qy	1321	PVDDAEVDVLQTKRSSRROSLELOKCEEILHKIVKTRFSWPPREPVTTRDEAEDYDVI	1380
Db	870	MCLADQKPLPRLPGLVLSGSTSDCLMVVQFLENFKVLGFDVNDVNLVLQGL	929	Db	1848	--INLSKQESFTSVKKPRKDDSDKOLALCSMILTEMETHEDAMPFLLPVNLKLVPGYKVI	1905
Qy	643	ADKGGFL-YLNRVLVILLQTLQDE-IADY---GELGMKLSEIPLTHSVSELVRLCLR	697	Qy	1381	THPMDFOIVONKSCGSYRSVQEFLLTMKQVPTNAEVN	1419
Db	930	LNIGSMGEVDLLVRLLSAAVCDPLGITGYAKTALGHEHLLNVGNVRDNVSEIILQIFM-	988	Db	1906	KKPMDFSTIREKSSGGQPNLETFPALDVLVDFDNCETFN	1944
Qy	698	RSDVQESGSDTDNDKDSAAEDNEVQDFLEKLETSEFFELTSEEKLIILTALCHRL	757	RESULT 15			
Db	989	-----BAHCGQTELSTKTKAFOAHTPAQKASVLAFLINEUA	1026	US-09-418-710-71			
Qy	758	MTYSVODHMETROQMSAELWKERLAVLKEENDKGAEKQKQKEMBAKNKENGKXVNG---	814	; Sequence 71, Application US/09418710			
Db	1027	CSKSVVSEIDKNIDYNSLRDKWVV---EGKLRKLRILIIHAKTKTKRTSGDIDLGEQ	1082	; Patent No. 6596482			
Qy	815	--LG-----KTRDKRIRVVFQVDTEADNMISAVKSRRLIAIAQAKEREIOE--REWK	864	; GENERAL INFORMATION:			
Db	1083	HELGTPTPKRRRRKGGSDYDDDDDDSD-----QGEDDEDEEDKEDQK	1129	; APPLICANT: Jones, Michael H.			
Qy	865	VKLERQAEERIRKHAAEK-----AFQEGIAKALVMRETPPTGTORNHR	911	; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR			
Db	1130	GKRTDICEDEGDAQASVEELEKQIEKLSKQOSQYRRKLFDASHLSLRVWFGPQYRRR	1189	; FILE REFERENCE: 06501-042001			
Qy	912	YMLFSDDEVPLFTE-----KGWVH-----DSIDYRFNHC	941	; CURRENT APPLICATION NUMBER: US/09/418,710			
Db	1190	YML-PRCGGIFVEGMESEGLEEBIAKEREKUKKAESVQIKBEMFETSGDSLNCNTDHC	1248	; PRIOR FILING DATE: 1999-10-15			
Qy	942	-----KDHT---VSGDEDEYCPRSK-----KANLGKNA-----	965	; PRIOR FILING DATE: 1998-04-17			
Db	1249	EQKEDLKEKDNITLFLQKPGSFKSLLEVAKMPPPEVMTPKPNAGANGCTLSYQNSG	1308	; PRIOR APPLICATION NUMBER: JP 9/310027			
Qy	966	--SMNTOHGTAEVAVE-----TTTTPKQGNLWFL-	993	; PRIOR FILING DATE: 1997-10-24			
Db	1309	KHSLGSVQSTATQSNVEKADSNLNFNTSGSGPGKFYSPLPNDQLLTKLTKEKRNQ--WFSL	1366	; PRIOR APPLICATION NUMBER: JP 9/116570			
Qy	994	-----CD-----	995	; NUMBER OF SEQ ID NOS: 73			
Db	1367	LPTPCDDTSLTHADMSTASLVTPQSPPSKSPSTPPAPLGSSAQNPVGLNPFALSPLOV	1426	; SOFTWARE: FastSeq for Windows Version 4.0			
Qy	996	-----	995	; SEQ ID NO 71			
Db	1427	KGVSMWMLQFCGWPTGVVTSNIPFTLSVPSLGLGLSEGNGNSFLTNSVASSKSESPV	1486	; LENGTH: 1876			
Qy	996	-----SQKELDELINCHLPQIGRES	1015	; TYPE: PRT			
Db	1487	PONEKATSQAQAAVEAVAKPVDPPSPKPIPEEMQFGWWRIIDPEDLKALLKVLHLRGIREK	1546	; ORGANISM: Homo sapiens			
Qy	1016	QKLERLEKRYQDIIHSIHLARKPNLGLKCDGNQOELLNFLRSLPIE---VATRLQKGGIGL	1072	US-09-418-710-71			
				Query Match 6.6%; Score 527.5; DB 4; Length 1876;			
				Best Local Similarity 19.5%; Pred. No. 8.3e-29;			
				Matches 2e8; Conservative 174; Mismatches 430; Indels 499; Gaps 44;			
Qy	415	OKSTGNSKSPKGLTPTKTKMQMTLLDMAGTKQ---KMTRAPRNSGGTPTSTSSKPHKHL	471	; APPLICANT: Jones, Michael H.			
Db	619	QAITKRCGRPR---NTERAKTKEVPKVKRGRGRPPKVKITEL-----LNKTDNRPLKXL	669	; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR			
Qy	472	PPAALHLIAYYKENDKDRKRSALSCVISTARLLSSEDRARLPEELRSVLQKRYELLEH	531	; FILE REFERENCE: 06501-042001			
Db	670	EA-----QETLNEEDKAKIAKSKKQKQKQVQGECLTTIQOARNKFKQETKSLKH	720	; CURRENT APPLICATION NUMBER: US/09/418,710			
Qy	532	KKRWASMSBEQKEYLKKRBEELKKLKEKAKEREKE-----M	570	; PRIOR FILING DATE: 1999-10-15			
Db	721	KB--AKKSKAKERKGTQKEKLKEKVKREKKEKEEYTKAPACKADTKLATQRR	778	; PRIOR FILING DATE: 1998-04-17			
Qy	571	LERLEKQ-----KRYEDQELTG-KNLPAFLVDTPPEG--LPNTLFGDVAMVVEFLSC	619	; PRIOR APPLICATION NUMBER: JP 9/310027			
				; PRIOR FILING DATE: 1997-10-24			
				; PRIOR APPLICATION NUMBER: JP 9/116570			
				; NUMBER OF SEQ ID NOS: 73			
				; SOFTWARE: FastSeq for Windows Version 4.0			
				; SEQ ID NO 71			
				; LENGTH: 1876			
				; TYPE: PRT			
				; ORGANISM: Homo sapiens			
				US-09-418-710-71			

Db 779 LEERQKQMI L EEMKKPTEDMCLTDHQPDPDFSRV---PGLTLP SGAFSDCLTIVEFLHS 835
Qy 620 YSGLLLPDAQYPTAVS-LMEALSADKGGFLYNRVLLVILLQTLLODETAEDYGE---L 674
Db 836 PGKVLGFDPAKDVPSLGVQLQEGLLCQDSLGEVDLLVRLKAAALHDPGFPSPYCSQSKKIL 895
Qy 675 GMLKSEIPLTLHSVSELVRLCLRRSDVQBESESGDNDKDSAAFPEDNEVDQEFLEKLET 734
Db 896 GEKVSIEPLTRDNVSEILRCFLMAYG-----VEPALCDRLRT 932
Qy 735 SEFFELTSEBKLOILTALCHRI---LMTYSVQDHMETRQMSAELW--KERLAVLKEEN 788
Db 933 QPFOAQPQKAAVLAFFPVHELNGSTLIINEIDKTLSESMYSYRKNKIIVEGRULRLK--- 989
Qy 789 DKRAEKQKREKMEAKNKENGKVENGLGKTDRKKRIKVFEPQVDTBAEDMISAVKRRLL 848
Db 990 -TVLAKRTGRSEVEM-----GRPECLGRRSSRIEETSGMEEEEEESIAAVPGR-- 1041
Qy 849 AIQAKKEREIOEREMKV-KLERQAEERIRKHAARAKAFQEGIAKAKLVMRTPIGTDR 907
Db 1042 ---GRDGEVDATASSIPELERQIEKLSK---QLFRRKKLLHSSQMLRAVSLGQDR 1092
Qy 908 NHRNYW-----LFSDEV-----PGLRIEK----- 926
Db 1093 YRRRYWVLPYLAGIFVEGTENLVPEVIKKTDSUKVAHAASLNPAFLFSMOMELAGSNT 1152
Qy 927 ----- 926
Db 1153 TASSPARASRLPKTKPGFMPREFKSPVRGQDSEQPAQLOPEAQLOHPAQPQLOLQ 1212
Qy 927 -----GWV-----HDSI----- 933
Db 1213 LQSHKGFLQEGSPLSGOSQHDLSQSAFLMSLSTQHSLSLSSSVLTDPSSPGKLDPA 1272
Qy 934 -----DYRENHCK-----DHTVSGDEDYCPRSKANKLGKA 965
Db 1273 SQPPEPEDEABSSPDLOAFWENISAQMPCNAAPTPPLAVSDEQ-----PTPSPQQLASSK 1329
Qy 966 SMN-----T 969
Db 1330 PMNRPSAANPCSPVQSSFTPLAGLAPKRRAGDPGEMPSQPTGLGQPKRRGRPPSKFKQM 1389
Qy 970 QHGTATEVAVETTPKQGNLWFLCDSQKELDBLNCILHPQIGRESQLEKRYQDII 1029
Db 1390 EORVLTLTAQAPVPPMCSGMMWIPDEM-LDAMLKALHPRGIREKALHKLNK-HRDFL 1447
Qy 1030 HSTHLARKNLGLKSCDGNQELNLFRLSDIEVATPL---OKG-----GLGY 1073
Db 1448 QEVCL--RPS-----ADPIPEPQLPAFQEGIMSWSPKETYETDLAV 1488
Qy 1074 VBETSFEARVIS-----LEKLKDFGECV----- 1097
Db 1489 LQWVELEQRVIMSDLOIRGWTCPSPDSTREDLAYCEHLSDSQEDITWPGREGIAPOR 1548
Qy 1098 -----IALQASVIKFLQGFMAPQO---KRRKIQSEDSA---KTEEVDEBK 1137
Db 1549 KTTNPLDLAVMLAALAEONVKKRYLREPLWPTHEWVLEKALLSTPNGAPEGTTTETSYE- 1607
Qy 1138 KMVEEAKVASALEKWKTAIREAQTFSRMHVLLGMLDACIKWMSAENAECKVCPKKGEDD 1197
Db 1608 -----ITPRIRIMRQTLQRCRSAAHVCLCLGHLERSIAWEKSVNKNVTCLVCRKGDND 1660
Qy 1198 KLILCDENKAFHLCLRLPALVEVPDGEWQCQACQAPATARRNSRGRNYTEESASESDED 1257
Db 1661 FLLLCDCGRGCHYCHRPMEAVPEGDFCTVC-----LAQOQVEGE 1702
Qy 1258 ESDEEEEBEEEEEDYEV---AGURLRPRKTIROKHVIPPAAASRRRPGKPHSTR 1314
Db 1703 FTQKPGFPRGQKRGKSYLNFSEGDGRRRRVLLKGRES---PAA-----GPYSEERL 1753
Qy 1315 SQKAPPVDDAEVDLVLQTKSSRRQSLQKCEBILHKIVKYRFSWFPREPVTRDAE 1374
Db 1754 SPSKRPL-----SMRNHSDLTFCFIIIMEMESHDAAWPFLEPVNPRLVS 1799

Qy 1375 DYYDVITHPMDFTQVQNKSCGSYRSVQBFLLTDMKQVFTNAEYVNCRGSHV 1425
Db 1800 GYRRIIKNPMDFSTMRELLRGYTSSEBFAADALLVFDNCQTFNEDDSEV 1850
RESULT 16
US-09-839-479-70
; Sequence 70, Application US/09839479
; Patent No. 672722
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042002
; CURRENT APPLICATION NUMBER: US/09/839,479
; CURRENT FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 70
; LENGTH: 1876
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-839-479-70

Query Match 6.6%; Score 527.5; DB 4; Length 1876;
Best Local Similarity 19.5%; Pred. No. 8.3e-29;
Matches 268; Conservative 174; Mismatches 430; Indels 499; Gaps 44;

Qy 415 QKSTGNSKSPKGLTKPTKMKQMTLLDMAKGTQ---KMTAPRNSGGTPTRTSSKPHKHL 471
Db 619 QAITGKRGPR---NTEKAKTEVPKVKRGRGPPKVKITEL-----LNKTDNRPLKKL 669
Qy 472 PPAALHLIAYYKENDREKRSALSCVTSKTRALLSSEDRARLPEELRSUVKRYELLEH 531
Db 670 EA-----QETLNEEDKAKIAKSKKMKQKQVQGECLTTIQGARNKPKQETSLKH 720
Qy 532 KKRWASMSBQRYKYLKKRBEELKKLKEKAKEREKE-----M 570
Db 721 K2--AKKSKABKRGKTKQEKLEKVKREKEKVKKEEVTKAKPACAKDTLATQR 778
Qy 571 LERLEKQ-----KRYEDQELTG-KNLPAFRLVDTPEG--LPNTLPGDVAMVVEFLSC 619
Db 779 LEERQKQMI L EEMKKPTEDMCLTDHQPDPDFSRV---PGLTLP SGAFSDCLTIVEFLHS 835
Qy 620 YSGLLLPDAQYPTAVS-LMEALSADKGGFLYNRVLLVILLQTLLODETAEDYGE---L 674
Db 836 PGKVLGFDPAKDVPSLGVQLQEGLLCQDSLGEVDLLVRLKAAALHDPGFPSPYCSQSKKIL 895
Qy 675 GMLKSEIPLTLHSVSELVRLCLRRSDVQBESESGDNDKDSAAFPEDNEVDQEFLEKLET 734
Db 896 GEKVSIEPLTRDNVSEILRCFLMAYG-----VEPALCDRLRT 932
Qy 735 SEFFELTSEBKLOILTALCHRI---LMTYSVQDHMETRQMSAELW--KERLAVLKEEN 788
Db 933 QPFOAQPQKAAVLAFFPVHELNGSTLIINEIDKTLSESMYSYRKNKIIVEGRULRLK--- 989
Qy 789 DKRAEKQKREKMEAKNKENGKVENGLGKTDRKKRIKVFEPQVDTBAEDMISAVKRRLL 848
Db 990 -TVLAKRTGRSEVEM-----GRPECLGRRSSRIEETSGMEEEEEESIAAVPGR-- 1041
Qy 849 AIQAKKEREIOEREMKV-KLERQAEERIRKHAARAKAFQEGIAKAKLVMRTPIGTDR 907
Db 1042 ---GRDGEVDATASSIPELERQIEKLSK---QLFRRKKLLHSSQMLRAVSLGQDR 1092
Qy 908 NHRNYW-----LFSDEV-----PGLRIEK----- 926

Db 1391 MEQRYLTQTAQVPPBEMCSGWWIPDPEDMLKALHPRGRGIREKALHKLHNK-HRDFL 1449
QY 1030 HSIHL--ARKPNLGLKSCDGNQE-LLNFLRSLDI EVATRIQKGLGGLGYVERTSEFEARVIS 1086
Db 1450 QEVCLRSADPIFEPRQLPAFOGIMSWSPKE-----KTYETDLAVLQWVELEQRVIM 1503
QY 1087 -----LEKLKDFGECV----- 1097
Db 1504 SDLQIRGWTCPSDSTREDLACEHLSDSQEDITWRGPGREGIAPQRTTNPLDLAVMRL 1563
QY 1098 IALQASVIKKFLOGFMAPKQKRKLQSDSAAKTEEVDEEKKMVE-----EAK 1144
Db 1564 AALEQNVKRYLRREPWP-----THEVLEKALLSTPNGAPEGTTTISYE 1609
QY 1145 VASALEKWKTAIREAQTFSRMHVLLGMLDACIKWMSAENARCKVCPKKGEDDKLILCDE 1204
Db 1610 ITPRIRIWRQTLQRCRSAHAHVCLCLGHLERSIAWKS VNKVTCLVCRKGDNDFFLLCDG 1669
QY 1205 CNKAFHLFCLRPALYEVDPGEWQCPACQATARRNSGRNRYTESASESDEDEEBE 1264
Db 1670 CDRGCHYCHRPKMEAVPEGDWFTVC-----LAQQVEGEFTQKPGF 1711
QY 1265 EEEEEEEDEYEV--AGLRLRPRKTIKRGKHSVIPPAARSGRPPGKPKPHSTRSQPKAPP 1321
Db 1712 PKRGQRKSGYSLNFGEGDGRRRVLLKGRS---PAA--GPR-----YSEERLSPSK-- 1759
QY 1322 VDDAEVDLVLQKRSRROSLELQCEELIKIVKRYRSPWPPEVTRDEADYDVIT 1381
Db 1760 -----RRRLSNHHSDLTCEIILMEMESHDAANFPXEPVNPRLVSGYRRIK 1808
QY 1382 HPMDFTQVQKSCGYSRVSQVEFLTDMKQVFTNAEYVNCRGSHV 1425
Db 1809 NPMDFSTMRELLRGGYTSSEEPADALLVFDNCQTFNEDDSEV 1852

RESULT 18

US-09-839-479-13
; Sequence 13, Application US/09839479
; Patent No. 6727222
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042002
; CURRENT APPLICATION NUMBER: US/09/839,479
; CURRENT FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 1878
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(1878)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-09-839-479-13

Query Match 6.5%; Score 521.5; DB 4; Length 1878;

Best Local Similarity 19.4%; Pred. No. 2.3e-28;

Matches 265; Conservative 179; Mismatches 437; Indels 483; Gaps 43;

QY 415 QKSTGNSKSPKGLKTPKTKMKQMTLLDMAKGTQ----KMTAPRNSGGTFRSSPKHKL 471

Db 619 QAITGRGRPR---NTEKATKEVPKVRGRGRPPKVKITEL-----LNKTDNRPLKKL 669

QY 472 PPAALHLIAYYKENDREDKRSALSCVISKTRALLASSEDRARLPEELRSLVQKRYELLEH 531
Db 670 EA-----QETLNEEDKAKIAKSKQMRQVQGECLTTIQGQARNKRKQETKSLKH 720
QY 532 KRWASMSBEQKEYLKKREELKKLKEKAKER---REKE----- 569
Db 721 KE--AKKKSXAEBKGTQKQKLEKVKREKKEKVQMKKEEVTAKAPACKADKTIALTOR 778
QY 570 MLERLEKQ-----KRYEDQELTG-KNLPAFLVDTPEG--LPNTLFGDVAWVVEFLS 618
Db 779 RLEERQKQKQWILEEMKPTEDMCLTDHQLPDPFSRV---PGLTLPFGASDCLTIVVEFLH 835
QY 619 CVSGLLLPDAQVPIITAVS-LMEALSADKGGFYLYNRVLVILLQTLLODRIADYGS---- 673
Db 836 SPFKVLGFPDPAKDVPSLGVQLQSGLLCQDGLSGEVQDLLVRLKKAALHDPGFPSPCQSLKI 895
QY 674 LGMKLSIEPLTHSVSELVRLCLRRSDVOEESGSDTDNKNDSAAAFEDNEVQDEFLEKLE 733
Db 896 LGEKVSEIPLTRDENVSEILRCFLMAYG-----VXPALCDRLR 932
QY 734 TGEFFELTSEEKLQILTALCHRI---LMTYSVQDHMETRQQMSAELW--KERLAVLKEE 787
Db 933 TOPFOAQPPQQAANLAFVPVHELNGSTLIINEIDKTLSSSYRKNKWIVEGLRLK-- 990
QY 788 NDKKRAEKQKREMEAKNKENGKVENGLGKTRKRIKVFEPQVDTAEADMTSAVKSRL 847
Db 991 --TVLAKRTGRSEVEM---GRPECLGRRSSRIMEETSGWEEEEEESIAAIVPGR-- 1042
QY 848 LAIAKKEREIEQEREMKV-KLERQABEEIRIKHKAEEKAFQEGIAKAKLVWRRTIGTD 906
Db 1043 ---GRDGEVDATASSIPELERQIEKLSK-----QLFFRKLLHSSQMLRAVSLGD 1092
QY 907 RNHNRYWLFSDVPGFLFI-----KGWVHDSID----- 934
Db 1093 YRYYRWL-PYLAGIFVGTGTEGNLVPEEVIKETDSLKVAHAASUNPALFSWKMLAGS 1151
QY 935 -----YRFNHCKDHTYSGDEDYCPRSK----- 957
Db 1152 NTTASSPARARSRLKTKPGFMQPRFKS-PVRGQDSEQFQAQLQPEALQHLVPAQPQPL 1210
QY 958 -----KANLGN-----ASMTQHGTEATEVAVETTPKQG-- 987
Db 1211 QLQLQSHKGFLQEGSPISLQSGQSHLSQSAFSLWSLQTSQSHSLSSSVLTPDSSPGKL 1270
QY 988 -----QNLWFLCDSQ----- 997
Db 1271 DPAPSQPEEPDEAESDPLOAFWNISAOMPNAAPTPPLAVSEDOPTSPQOLASS 1330
QY 998 ----- 997
Db 1331 KPMNRPSAANPCSPVQFSSTPLAGLAPKERRAGDPGEMPOSPTGLQPKRRGRPPSPKFKQ 1390
QY 998 -----KELDELNCLHPQIGIRSQKLERLEKRYODII 1029
Db 1391 MEQRYLTQTAQVPPBEMCSGWWIPDPEDMLKALHPRGIREKALHKLHNK-HRDFL 1449
QY 1030 HSIHL--ARKPNLGLKSCDGNQE-LLNFLRSLDI EVATRIQKGLGGLGYVEETSEFEARVIS 1086
Db 1450 QEVCLRSADPIFEPRQLPAFOGIMSWSPKE-----KTYETDLAVLQWVELEQRVIM 1503
QY 1087 -----LEKLKDFGECV----- 1097
Db 1504 SDLQIRGWTCPSDSTREDLACEHLSDSQEDITWRGPGREGIAPQRTTNPLDLAVMRL 1563
QY 1098 IALQASVIKKFLOGFMAPKQKRKLQSDSAAKTEEVDEEKKMVE-----EAK 1144
Db 1564 AALEQNVKRYLRREPWP-----THEVLEKALLSTPNGAPEGTTTISYE 1609
QY 1145 VASALEKWKTAIREAQTFSRMHVLLGMLDACIKWMSAENARCKVCPKKGEDDKLILCDE 1204
Db 1610 ITPRIRIWRQTLQRCRSAHAHVCLCLGHLERSIAWKS VNKVTCLVCRKGDNDFFLLCDG 1669
QY 1205 CNKAFHLFCLRPALYEVDPGEWQCPACQATARRNSGRNRYTESASESDEDEEBE 1264

Db	1670	CDRGCHYICHRPKMEAVPEGDWDFCTVC-----LAQQVEGEFTQKPGF	1711
Qy	1265	EEEEEEEDYEV---AGLRIPKRTTRGKHSVTPPAARSGRRPGKXPHSTRRSQKAPP	1321
Db	1712	PKRQKRGKGYSLNFSFGDGRRRRLVLKGRES---PAA--GPR-----YSEERLSPSK--	1759
Qy	1222	VDDAEVDVLVLOTGRSSRRQSLELQKCEILHKIVKRFSPWPPEPVTTRDEAEDYDVIT	1381
Db	1760	-----RRRLSRNRNHSDDLTFCEIILMEMESHDAAPFEPYPNRLVSGYRRIK	1808
Qy	1382	HPMPDQTQVQKSCGSRVSRVQEFLLTMKQVFTNAEVVNCGRSHV	1425
Db	1809	NPMDFSTMRELLRGYTSSEEFADALLVPDNCQTNEDDSEV	1852

```

RESULT 19
US-09-418-710-72
; Sequence 72, Application US/09418710
; Patent No. 6596482
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042001
; CURRENT APPLICATION NUMBER: US/09/418,710
; CURRENT FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 72
; LENGTH: 1969
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-418-710-72

```

Query Match	6.4%;	Score 508.5;	DB 4;	Length 1969;	
Best Local Similarity	18.0%;	Pred. No. 2.le-27;			
Matches 353;	Conservative 246;	Mismatches 601;	Indels 763;	Gaps 68;	
Qy	8	KPFFLVNPLPCEPPF-FTHPH-	-----TQAFRTREERYEARLERYSERIWTCKST	55	
Db	189	KPLSLVNQAKKETYTKMLIVSPDVLKAGNKNTSESSLTSELSKEQY	-----KQA	241	
Qy	56	GSSQLTKHKAWESEQEVAELLK	-----EEFPAMYEKVLVLEMVHIN	TA 97	
Db	242	PPSOLKTKQESSKSLKVKVIAALSNPKNTSSSPA-HPKOTLENNHPNPLTNALGNHPNG	300		
Qy	98	SLEKLVDTAWLEIMTKYAVGECDEFV	-----	124	
Db	301	VIQSVIQEAPLALTKTKMKQSKINENTAAASSTPFSVPVNLSTSGRRTPGNQTPMPSAS	360		
Qy	125	-----GKEKML--KVIVKI	-----HPL	-----EKVDE 145	
Db	361	PILHSQGEKAVSNVNNVPVKTQHHSHPAKSLVQFRGTDSDIPSSKDSDESDNEDEE	420		
Qy	146	EATEKKSOGACDSPSDKENSQJTAQHQKKEVTVKDEGRRESINDPARRSPRKLPTSL	205		
Db	421	EBEDEDEDDSDSOESDSNSE--SDTEGSEEDDDDKQOESDSD	-----	465	
Qy	206	KKGERKWAPPFLPHKVDVKLQNEDKIISNVPADSLINTERPPPKKEIVRYEIRHNALRAAG	265		
Db	466	TEGE	-----KTSMKLNKTTTS--SKSPSNLSTGHSTERNLHIAK	-----	501
Qy	266	TGENAPWVVEDELVKKYSLPSPKSFDFLLDPKYMTLNPSTTKRKNVTSQDPDRKPSKSK-TD	324		
Db	502	----APGSAPAAALCSQSQSA	-----	-----FLGTSSSTL--TSSPHGTSKRRRVTD	543
Qy	325	NSLSLSSPLNPKLWCHVHLKKSLSGSPULKVKNKSNKSP	-----EEHLEMMKWGSPN	-----	376


```
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 11433
; LENGTH: 2733
; TYPE: PRT
; ORGANISM: Human
; US-09-949-016-11433

Query Match          3.8%; Score 304; DB 4; Length 2733;
Best Local Similarity 19.3%; Pred. No. 1.8e-12;
Matches 351; Conservative 280; Mismatches 606; Indels 580; Gaps 87;

QY 30 EAPRTREYEARELYSERIWTCKSTG-----SSQLTHKEAMEEEOEVAELKEEPPA-- 82
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
505 ETERGEVEEDKENKEISEKCVTSKQOEIETLYLKQITISEKEV--ELQHIRKDLKEELAAEE 562
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 83 WYKLVLEM-----VHNTASLEKLV-----DTAWLEIMTKYAV 116
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

563 QFOALVKMNQTLQDKNQIDLLQAEISENOAIIQKLITNTDASDGDVAL-VKETVVI 621
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 117 GECDEEVG-----KEKMLKVKIVKIHPLKLEKVDDEATEKKS--DGACDPSDDKEN 165
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

622 SPCC--TGSSEHWKPELEKILALEKEKEQLQKQLEALTSRKALKKAQAEKERHLEE 678
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

166 SSQIAODHOKETVTVKXEDGRRSINDRARRSPKLTSLKGGKRWAPKPLPHKYDVK 225
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

679 LKQKDDYNRLQEQPDEQSKENENIGDQLRQLQIQVRESI-----DGK 721
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

226 LONED--KIISNVA-----DSLIRTPPNKSIIVRYFIRHNALRAG 265
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

722 LPSTQDQESCSSTPGLPEPLFKATEQHTQPVLESNLCPDWPSHSDA-----SALQGG 775
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

266 TG-ENAPWVVEDELVKYSLPSKFSDFLDPPKYMTLNPSTKRNKTSPPDRKPSKSKTD 324
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

776 TSAQITKAQLKEIAEKAEKVELEKVS-----TTSELTKSEEVFQLOEQINQGLE 826
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

325 NSLSPLNPKLVCHVH---LKKSLGSPLVKVNKSNKSPSEHLEEMKMSPNK--- 377
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

827 IESLKTVSHE---AEVHAESLQKLESSOLOIAGLEHLRELQPLDELQKLISKEEDVS 883
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

378 -----LHTNFHIPKGPAPKPKGHSK----- 400
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

884 YLSGQLSEKAAATKIQTETIEQEDLIKALHTQLEMOAK-----EHDRIKQLQVEL 935
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

401 -PLKAGRSKGIINGOKSTGNSK-----SPKGLTKPTKQKQMTLLDMAGKT-QQWTR 452
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

936 CEMKQPEEIGESRAKQIQRLQAALISRKALKENKSLQBELSL---ARGTIERLTK 992
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

453 APNSGGTPTRTSSKPHKPLPAAHLHIAVYKENKDR-----EDKRSALSCVISTK 502
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

993 SLADVESQVSAQNK-----KDTVLGRLLQBERDRLITEMDRSLLENQSLSSCSLKL 1048
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

503 ARLLSSDDRARLPPEELRSL-----VQKRYELL-----BHKRWK----- 536
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

1049 ALEGLTDEKLVKEITESLKSSIAESTEWQEKHEKELQKEYEILLQSYENVNSNEARIQH 1108
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

537 ----SMSEORKEYLKKRRELKKLKAKERKEKEMLERLEKQRY----- 580
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

1109 VVEAVRQEQELYGLKRLSTEANKKETEKQLQEAQEMEBMKQKRFKAKSKOQKILELEE 1168
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

581 EDQELGKNLPAFLVDTPTEGLNTLFGDVMVVEFLSCVSGILLLLDAQYPIITAVSLMEA 640
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

1169 ENDRLRAEVHPA-----GDTAK-----ECMETLLSSNA-----SMKEE 1201
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

641 LSADKGGFLYLNRLVILLQTLQDEIADYGLGMLKLSIPLTLHSVSELVRLCLRRSD 700
```

US-09-949-016-6507
; Sequence 6507, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6507
; LENGTH: 3259
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-6507

Query Match 3.8%; Score 303; DB 4; Length 3259;
Best Local Similarity 19.3%; Pred. No. 2.7e-12;
Matches 348; Conservative 285; Mismatches 614; Indels 560; Gaps 85;
QY 30 EAFRTREYEARLERYSERIWTCKSTG-----SSQLTHKEAWBEEQVAAELLKEFPFA-- 82
DB 1031 ETERGEVEDKENKEYSEKCVTSKQEIYILKQITISEKEV--ELQHIRKDLKEELAABE 1088
QY 83 WTEKLVLEM-----VHNTASLEKLV-----DTAWLEIMTKYAV 116
DB 1089 QFQAQVKOMQNTLODKTNQIDLOLQAEISENQAIQKLTITSNDASGDSVAL-VKETVVI 1147
QY 117 GEEDCFEVG-----KEKMLVKVIKIHLEKVDDEATEKKS--DCACDPSDDKEN 165
DB 1148 SPPC---TGSSEHWPKEPEKILALEKEKEQKLOKQALTSRKAILKKAQAKERHLREE 1204
QY 166 SSQIAQDHQKXTVVKEDGREGRESINDRARRSPRLKPTLSLKXGERKWAAPKPLPHKYDVK 225
DB 1205 LKQKDDYNRLQEQDFEQSKENENIGDLRLQRLQIQVRESI-----DGK 1247
QY 226 IQNED--KLIISNVA-----DSLIRTERPNKEIVYFIRHNALRAG 265
DB 1248 LPSTDQCESCSPTGLEELFKATEQHHTQPVLSENLCPDWPFSHGSA-----SALQGG 1301
QY 266 TG-ENAPWVVEDELVKYSLPKSFDFLLDPYKYMTLPNTSPKXONTGSDPKSPKSKTD 324
DB 1302 TSAQIKAQKLEIAEKVLELVSS-----TTSLETKSBEVFOLOQINKQGLE 1352
QY 325 NSSLSPLNPKLWCHVH---LKKSLSGSPKVNKNKSKSPPEHLEEMMMKMSPNK----- 377
DB 1353 IESLKTVSHE--AEVHAESLQKLESSQIAGLEHLRELQPKLDELQKLISKKEEDVS 1409
QY 378 -----LHNFHPIPKGPPAKPKGKHSK----- 400
DB 1410 YLSGQLSEKAALTQITETIEQEDLIKALHTQLEMOAK-----EHDRIKQLOVEL 1461
QY 401 -PLKAKGRSGKILNGOKSTGNSK-----SPKGLKTPKTKMOMTLDMAGKT-QKMTR 452
DB 1462 CEWKQKPEIGESRAKQIQKQLQALISREALKENKSLQBELSL---ARGTTERLTK 1518
QY 453 APRNSGGTPTSTSKPHKHLPPAALHLIAYYKENKDR-----EDKRSALSCVISKT 502
DB 1519 SLADVESQVSAQKE---KDTVLGRLLALQBERDKLITEMDRSLLENQSLSSCSBLKL 1574
QY 503 ARLLSESDRARPPEELRSL-----VQKRYELL-----EKKRWA----- 536
DB 1575 ALEGUTEDKLVKETIESLSSKSIABSTEWQEKHLEKQYELLQSYENVSNEAPRIQH 1634
QY 537 ---SMSEQRKEYLKKKREELKKLEKAKERREKEMBLERLEKQRY----- 580

DB 1635 VVEAVRQEQELYGKLRSTEAANKETEQLOQAEQEMEEMKMKFAKSKQOKILELBE 1694
QY 581 EDQELTGKGLPAFRLVDTPEGLPNTLFGDVAMVVEFLSCYSGLLLPDAQYPTAVSLMEA 640
DB 1695 ENDRLRAEVHPA-----GDTAK-----ECMETLLSSNA-----SMKEE 1727
QY 641 LSADKGGFLYLRNLVILLQTLLODEIARDYCELGKMLSEIPLTLHVSSELVRLCLRRSD 700
DB 1728 LERVQNEYETLSKKFQSLMSE--KDSLEEVODLKHQIED-----NVSK----- 1769
QY 701 VQESSEGSDDNDKDSAAAFEDNE-VODEFLEKLETSEFFELTSEBKLIQILTALCHRLMT 759
DB 1770 -QANLEATEKHNDQNTVTEEGTQSPGETEEQDSLSMSTRPTCSSESVPSAKSANPAVSKD 1828
QY 760 YSVQDHMETROQMSABLWKLKRLAVLKEENDKKR-----AEKOK----- 797
DB 1829 FSSHDEINNYLQOIDQL--KERIAGLEEBKQKNKFEFSQTLSENEKNTLLSQISTKDGELKML 1887
QY 798 RKEMEAKNKNGKVENGLGKTDKRIIVKFPQVDTAEADMISAVKSRLL----- 848
DB 1888 QEEVTMMLLNQIOBEL-----SRVTKLETAEKDDL-----EERLMNQLAELNGS 1936
QY 849 -----AIOAKKEREIOEREM-----KVKLERQAAEEERIRKHAATAA-----KAFQEGIAK 893
DB 1937 IGNYCQDVTDAQIKNELLESEMKLKKCVSELEEEKQLVKEKTKVESEIRKEYLEKIQG 1996
QY 894 AKLVMRRPTIGTDNRHNRWYLFSDVPGLFIE-----KQWVHDSIDYRPNHHCKDHTVSG 948
DB 1997 A-----QKEP--GNTKSH-----AKELQELLKEQOEVKQLOKDCIRYQEKISALERTVKA 2044
QY 949 DEDYCPRSKK--ANLGKNASMTQHG-----TATEVAVENTTPKQGNLWFLCDSQKE 999
DB 2045 LEFVOTESQKOLEITKENLAQVHRKKAQABLASFVKLLDDT---QSEAAARVLADNLKL 2101
QY 1000 LDELINCLHPQGIRES---SOLKER---LEKRYQDIHSHILARKPNLGLKSCDGNQELN 1053
DB 2102 KKEEL-----QSNKESVKSQMKQDEDLERRLEQ-AEEKHLKEKNM-----QEKLD 2146
QY 1054 FURSLI-----EVATRLQKG--LGYVEE-----TSEFPAFVISEKLKDFGECVIA 1099
DB 2147 ALRREKVLHEETIGIQTLLNKKQKEVQOQLQENLDSVTVQLAAFTKSMSSLDQDRDRI 2206
QY 1100 LQASVVKFLQGFMAPKOKRKLQESDAKTEVEDEEKKMVEAKV-ASALEK---WKT 1154
DB 2207 BAKKWERKPSDAIQSKEBEIRLKEDNCVILKQDLQMSHMEELKINTSRLEHDKQIWES 2266
QY 1155 AIREAQTFSRM-----HVL-----LGMLDACIK-----WD 1179
DB 2267 ---KAQTEVOLQKVCDDTLQGENKELLSQLEBTRHLHYSSQNELAKLESELKSLQDLTD 2323
QY 1180 MSAENARCKVCPKGEDDKLILCDECNKAFHLCRLPALYEVDPDGEWQCPACQATARN 1239
DB 2324 LNSLSLEKCK--EOKNLEGIIRQEQAD-----IQNSKFSYEQLETDLOASRELTSLRH 2374
QY 1240 SGRNVYTE-----SASEDS-----EDDESDE-----EEEEEE----- 1273
DB 2375 EE-INNKQKIISLSGKEEALQVAIAELRQOHDKEIKELNLSQEEENIVLEENKK 2433
QY 1274 DYEAVGLRLRPRTKIRGHSHVIPPAAASGRPRPKGPHSTRSOPKAPPVDDDAEVBELV-- 1331
DB 2434 AVDTKNQLMETLTKTIK-KENI-----QOK-----AQLDSFVKS 2465
QY 1332 ---LQTKS---SRQSLQELQCKEELHVKIVYRFSWFPREPVTRDEA-----EDYD 1378
DB 2466 MSSLQNDRIIVGDYQOLEERHLSILEK-----DQLIQEAAENKNKLEIRG 2514
QY 1379 VLTHPMDFTQVQNKSCGYSRVQBFLLTDMKQFTNAEVYNCRGSHVLSVMWVKTEQLVV 1438
DB 2515 LRSHMDDLNSENAKLDA-----ELIQTREDLNQVIT-----IKDSQ-QKQ 2553
QY 1439 LLHKLHPGHPYVRRKPKFPDLAEDGDSPEAVQSGRDEDRRRREABIQEWLQDTS-- 1496

Db 2554 LLEVQLOONKELNRYAKLEEKKESE-----EANEDLRRSFNA-LQEEKQDLSKE 2603

QY 1497 LYSAKIN 1503

Db 2604 IESLKVS 2610

RESULT 25

US-08-875-435B-3

; Sequence 3, Application US/08875435B

; Patent No. 6593304

GENERAL INFORMATION:

; APPLICANT: Hasegawa, Kazuhide

; APPLICANT: Arakawa, Emi

; APPLICANT: Oda, Shoji

; APPLICANT: Matsuda, Yuzuru

; APPLICANT: Takahashi, Katsuhito

; APPLICANT: Sugahara, Michihiro

; APPLICANT: Ishiyama, Haruo

; TITLE OF INVENTION: RECOMBINANT DNA COMPRISING DNA CODING

; TITLE OF INVENTION: FOR MYOSIN HEAVY CHAIN SM1 ISOFORM PROTEIN INSERTED INTO

; TITLE OF INVENTION: VECTOR DNA, MICROORGANISM CARRYING THE RECOMBINANT DNA, AND

; TITLE OF INVENTION: AN AGENT FOR TREATMENT OF ARTERIOSCLEROSIS COMPRISING THE

; TITLE OF INVENTION: RECOMBINANT DNA

; FILE REFERENCE: 07898-013001

; CURRENT APPLICATION NUMBER: US/08/875,435B

; CURRENT FILING DATE: 1997-07-25

; PRIOR APPLICATION NUMBER: PCT/JP96/00134

; PRIOR FILING DATE: 1996-01-25

; NUMBER OF SEQ ID NOS: 5

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 3

; LENGTH: 1972

; TYPE: PRT

; ORGANISM: Mus musculus

US-08-875-435B-3

Query Match 3.8%; Score 300.5; DB 4; Length 1972;

Best Local Similarity 19.4%; Pred. No. 2e-12;

Matches 280; Conservative 233; Mismatches 533; Indels 397; Gaps 60;

QY 12 LVNPLGPEEPFFIPIHTQAEAFRT-----REEVEARLERYSERIWTCKSTGGSQLTHKEA 65

Db 750 MIKALDELNLYRIGOSKIFPTFVLAHLEERDLKI---TDVIMAFQAMCGYLARKAF 806

QY 66 WEEQVFAELLKEFPANWEKLVLEWHNTASLEKLVDTAWLEIMTK-----YAVGEBC 120

Db 807 TKRQOQLT-----AMKVIQRNCAAYLKLNRNQMWRLLFTKVKPLLOVTRQEE 852

QY 121 DFEVGKEKMLKV-----KIVKTHPLEKVDDEATEKKS-----DGACDSPSDKENSQ 168

Db 853 EMQAKSEEMQKITERQQAETELKELEQKHTQLAEKTLLOEQLOAETELYAESEEMVR 912

QY 169 IAQDHQKKTWKEDSGRRSINDRARRSPKLTSLKGERKWAHPKPLPHKYDVLQN 228

Db 913 LAAKQOELEILHEMARLEEDRRQ-----LQAERK-----KMAQQMLDLEQL 959

QY 229 EDKIISNPADSLIRTERPNKEIVFYFIRHNALRAGTGENAPWVVEDELVKYSLPSKF 288

Db 960 EE-----EBAARQKLQ---LEKVTAEAKTKKLEDDILVMDQNSKL 997

QY 289 S---DFLLDPYKVTMLNPSTKRKNTGSPDRKPSKSTONSSLSPLNPKLWCHVHLKKS 345

Db 998 SKERKLLERVSDLTTLNLAEEKA-----KNITLKSKHESMISEL-----EVLKKE 1046

QY 346 LSG-----SPLKVNKSNKSPBEHLEEMMMOMSPNKLHTNFHIPKGPAPKPGKSHDKP 401

Db 1047 EKSRLQELKRLKLEGDASDPHEQIADLQAIAELKM---QLAKKEEELQAALARLDEE 1102

QY 402 LKAKGRS-KGI-----INGQKSTGNSKSPK-----GLKTPKTKMQMTLLD 442

Db 1103 IAQKNALKKIRELEGHISDLQEDLDSERAARNKAEKQKRDGLGEELAKTLELED--TLD 1160

RESULT 26

US-09-764-176-7

; Sequence 7, Application US/09764176

; Patent No. 6809189

QY 443 MAKQTQOMTRAPRNSGGT-----PRTSSKPHKHLPPAALHLIAYYKKNKOREDKRSL 495

Db 1161 -STATQOELRAKREQEVTVLKALDEETRSHEAQVOEMRQKHTQAVAEELTEQLEQPKRA- 1218

QY 496 SCVISKTARLLSSEDRARLPEELRSIVOKRYELLEHHKKWASMSSEQRKEYLKKKEELK 555

Db 1219 KANLDKSKOTLEKEN-ADLAGELRVLGQAKQE-VEHKKK-----KLEVQLODLQ 1265

QY 556 KKLKAKAKERREKEMLEKQRYEDQBELTKGNLPAPFLRVDTPEGLPNTLFGDVAWVVE 615

Db 1266 SKCSD--GERARAEISDKVHKLQN-EVESVTG-----MLNEAEKAIKLAQDVASLGS 1315

QY 616 FLSCYSGLLLPDAQYPIITAVSLMEALSADKGGFLYLNRLVILLQTLLODEAED----- 670

Db 1316 QLQDTQELLQEEETROKLVNSTTKLRQLEDERNS-----LQDLDEEMAEKQ 1360

QY 671 -----YGELGMKLSIPLTLHSVSELVRLCLR-RSDVQSESESGSDTDDKNSAAFE--- 720

Db 1361 NLERHVSTUNIQLSDSKKKLQDPASTIEVNEGKGLQKEMEGLSQOYEEKAAAYDKLEK 1420

QY 721 -DNEVQDEF-----LEKLETSEFFELTSEELKILQILTALCHRLIMTYSVQ 763

Db 1421 TKNRLQOELDLVLDLNDQRLVSNLEK-KQKKFDOLLAEEK-----NISSKYA-- 1468

QY 764 DHMETROQMSAEL-WKERLAV-----LKEENDKKRAEKQKREKEMAKKNKENVEN 813

Db 1469 ---DERDRAEAEBAREKETKALSARALAEBALEAKELELERNMLKAEMEDLVSSKDDVGK 1525

QY 814 GLGKTDRKKRIVKFPPOVD-----TEADMTISAVKSRRL-----LATQAKKERIOER 861

Db 1526 NVHELEKSKRAL--ETQMEEMKTQLEESDDQVATEDAKLRLEVNQNALKGOFERDQAR 1583

QY 862 ----EMKVKLERQ-----AEERIRKHAAAEKAFQEGIAKAKLVNRRPTI--GTDNRN 908

Db 1584 DEQNEKRRQORQORLHEYTELEDERKQALAAAKKLEGDLKOLELQADSIAKRE-- 1641

QY 909 HNRWLFSDPEVPLFTIEKGWVHDSIDYRFNHHCKOHTVSGDBDYC---PRSKANLGKNA 965

Db 1642 -----EAIQLRKLQAKMD-----FORELDDARASRDEIFATSKENKKAKSLEAD 1688

QY 966 SNTQHGAT-----EVAVETTPKQGNLWFLCDSQKELDELLNCLHPQGI 1012

Db 1689 LMQLQEDLAAERARKQADLEKEBELASLSGRNT--LQDEKERLE-----A 1736

QY 1013 RESOLKERLEKRYQDIHSIHILARKPNLGLKCDGNOELLNFRSLIEVATRLQKGGGLG 1072

Db 1737 RIALEEELEEE-----QGNWEAM-----SDRVKAT----- 1763

QY 1073 YVEETSEFARVISLEKLDGECVIALQASVIKFLQGFMAPKQKRRKLQSEDSAKTEE 1132

Db 1764 -----LQAEQLSNE-----LATERSTAQK---NESARQQLERQNKELSKLOE 1803

QY 1133 VDDEKK-----MVEEAKVASALEKWKTAIRBAQTFSRMHVLLGLMDACIKWDMSEAENA 1185

Db 1804 VEGAVKAKLKSTVAALAKIAQLEEQVEQAEKQAATK----- 1842

QY 1186 RCKVCPKGEDDKLILCDECNKAFHLFCLRPALYEVDPGECWQCPACQAPATARRNSGRNY 1245

Db 1843 -----SLKQKDKK-----LKEVLQVEDERKMAEQKEAQKGNKTVKQL 1882

QY 1246 TE--ESASDSEDDSEDEEEEEEEEEEEEDYEAGLRRLPRKT--IRGKHSVIPPAAARS 1301

Db 1883 KQLEEAEBESOCINARRKLORELDDEATESNEAMGREYNALKSLRRGNASVFSRRA 1942

QY 1302 GRR 1304

Db 1943 GGR 1945

```

; GENERAL INFORMATION:
; APPLICANT: NOTEBORN, Mathieu Hubertus Maria
; APPLICANT: DANEN-VAN COORSCHOT, Astrid Adriana Anna Maria
; APPLICANT: ROHN, Jennifer Leigh
; APPLICANT: WEISS, Bertram
; TITLE OF INVENTION: APOPTIN-ASSOCIATING PROTEIN
; FILE REFERENCE: 4725US
; CURRENT APPLICATION NUMBER: US/09/764,176
; CURRENT FILING DATE: 2001-01-17
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 7
; LENGTH: 1400
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: misc feature
; OTHER INFORMATION: Description of Sequence: Amino acid sequence deduced from the nucleotide sequence of the cDNA
; OTHER INFORMATION: leuc acid sequence of AAP-
US-09-764-176-7

Query Match      3.8%; Score 299; DB 4; Length 1400;
Best Local Similarity 19.4%; Pred. No. 1.5e-12;
Matches 308; Conservative 245; Mismatches 551; Indels 480; Gaps 78;

QY 3 PLIG--RKPP-----LVNPLP-----GEPFTIHTQEAFF-----TREEYEARLERY 45
DB 25 PLDLPPELPPPELVRQAPPDVGNGEVPKELVHLHLMKPKIKGKSVTADRWEKYLKY 84
QY 46 SERIWTCKSTGSSQLTHKEAWEEGEVALL-----KEEFPWYKLVLEVMVHENTASL 99
DB 85 -----LCEQPDNLKFKNIINEEDADTMRLOPIGRDKGLMYWYQ---LDQHNVRMYI 136
QY 100 EKLVD---TAWLEIMTKYAVGECDFEVGKERKMLKVIKVIHPLBKVDEEATEKSGDGC 156
DB 137 EQDDQDQSSW-----KCIVRNRNELAETLALLKAQ-IDPVLLKNSSQDQNSRR 184
QY 157 DSPSSDKENSSQIAQDHQKKEITVVKEDGRRRSINDRRARRSPKPLTSLKGERKWAPPK 216
DB 185 ESPSLEDEET-----KKEETPKQEEKSEKKEEQPMDL----- 221
QY 217 FLPHKYDVKLQWEDKIISNVPADSLIRTPPNKEIVRYFIRHNALRAGTGENAPWVVED 276
DB 222 -----ENRSTANVSETTVKKEDEKELV-----LPVIVKL 254
QY 277 ELVKYSLPSKSFDFLLDPYKMTLNPSTKRNKTSPPDRKPSKSKTNSLSPLNPKL 336
DB 255 E---KPLPE-----NEEKIIKEESDSFKNVKPIKVEVKECRADPKOTKS----- 297
QY 337 WCHVHLKSLGSLPKVKVKNKSKS-----PEHLEEMKQMSPN-----KLHTN 381
DB 298 -----SMEKPAQEPERIEFGNIRKSHETKSTETKLNQDQAKPLKKREIKLSDD 353
QY 382 FHPKKGPP-----PAKPKGKHSK-----PLKAGK-RSKGILNGOKS-----TCN 420
DB 354 FDSVPKGLPKSVTPFKELKDEIKQEEETCKRISTITLALGHEGQLNGEVSDSERVAPN 413
QY 421 SKSPKKGKLTPTK-----MKQMTLLDWAQTKQMTAPRNSGGTPTTSKPHKHLPPA 474
DB 414 FKTEPIETKFYETKEESYSPSKDRNIITEGNGTESL-----NSVITSMKTGELEKETAPL 468
QY 475 ALHLAYYKENDREKRALSC--VISTARLLSSEDPARLPEELRSIVQRYELLEHK 532
DB 469 -----RKDADSSIVLEIHSQAKIEBPD-----PPMETSLDSS-----EMA 506
QY 533 KRWASMSERQKYLEYKKEEELKKLKEKAKERREK--EMLRERLEKQKRYEQELTGKYL 590
DB 507 KDLSSKTLSSSTESCTMKGEESPKTK---KDKRPPILECLEKLEKSK-----TFLDK 557
QY 591 PAPRLVDTPEGLP-NTLFGDVAWVVEFLSCYSGLLLPDAQYPTAVSMALSAADKGGFL 649
DB 558 DAQRLSPIDEEVPKSTLESEKPSPE-----AAETSPSPSNIIDHCEKLAISEK----- 604
```

RESULT 27

```

US-09-538-092-1131
; Sequence 1131, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
; APPLICANT: Giot, Loic
; APPLICANT: Mansfield, Traci A.
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538,092
; CURRENT FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
```


RESULT 29

;; TITLE OF INVENTION: EXPRESSION PRODUCTS THEREFROM, AND USES OF SUCH GENES,
;; TITLE OF INVENTION: PORTIONS AND PRODUCTS

;; NUMBER OF SEQUENCES: 47

;; CORRESPONDENCE ADDRESS:

;; ADDRESS: Curtis, Morris & Safford, P.C.

;; STREET: 530 Fifth Avenue

;; CITY: New York

;; STATE: New York

;; COUNTRY: U.S.

;; ZIP: 10036

;; COMPUTER READABLE FORM:

;; MEDIUM TYPE: Floppy disk

;; OPERATING SYSTEM: IBM PC compatible

;; SOFTWARE: PC-DOS/MS-DOS

;; CURRENT APPLICATION DATA:

;; PATENT IN RELEASE: 1.0, Version #1.30

;; APPLICATION NUMBER: US/08/714,741

;; FILING DATE: 16-SEP-1996

;; CLASSIFICATION: 435

;; ATTORNEY/AGENT INFORMATION:

;; NAME: Frommer Esq., William S.

;; REGISTRATION NUMBER: 25,506

;; REFERENCE/DOCKET NUMBER: 454312-2460

;; TELECOMMUNICATION INFORMATION:

;; TELEPHONE: (212) 840-3333

;; TELEFAX: (212) 840-0712

;; INFORMATION FOR SEQ ID NO: 32:

;; SEQUENCE CHARACTERISTICS:

;; LENGTH: 8991 amino acids

;; TYPE: amino acid

;; STRANDEDNESS: single

;; TOPOLOGY: linear

;; MOLECULE TYPE: amino acid

;; US-08-714-741-32

Query Match 3.7%; Score 294.5; DB 4; Length 8991;

Best Local Similarity 18.0%; Pred. No. 4.6e-11; Indels 601; Gaps 74;

Matches 334; Conservative 284; Mismatches 655;

24 TIPDQEA-----FRTREYEALRYSERIWTCKSTGSSQLTHKEAMWEEQVEAEELKE 78

1395 TVVRAEAPVASQSKAEKDYDA- AKGDAEN--AKKALEBAKRAQKXEDDQKTEKAKX 1451

79 EFPA-----WTEKLVLEWHINTASLEK-----LVDTALEIMTKYAVGEEDCFEVG 125

1452 DXQASEAEQKANLXYQLLQKYVSSDGKKKEXEXXADAAKEIEXLXADLXIXOENX 1511

126 KEKMLKVKIVKI-----HPLEKVDEEA-----TEKSDGACDSPSSDKENSSQ 168

1512 KKNMILTSASVAILGAGLVASQPTVVRAEAPVASQSKAEKDYDAAEKSAEEDLEE 1571

169 IAQDHQK--ETVVKDEGRESINDRARRSPKLTSLKGERKWAPKFLPHRYDYKL 226

1572 -AEAQRKYDEQKSEENEKETEASERQQAATLKYHLESX-----FLNYFDNHR 1623

227 QNEDKIISNPADSLI-----RTERPPNKEIVRYFIHNLALRAGTGENAPWVVEDELVKY 282

1624 MNKKMILTSASVAILGAGLVASPTT--VVR-----AEAPVASQSKAEKDY 1669

283 SLPSKPSFLDLPYKYMTLNPSTKRNKNTGSPORPKSKGSKTNSLSPLNPKLVCHVL 342

1670 DTAQRDAE---NAKALEBAKRAQAEKYADYQRIEKAETHASLEQOEANKDY-QLKL 1725

343 KKSLSGSPLVKXNSKNKSPPEHLEEMKQMS-----NKLHNTNPHI PKKGPAPKPKGH 397

1726 KKYLDG-----RNLSNSVLKXMEEAEEKKDEKPAEFNKIRREIWP--NQPELEMAR 1778

398 SDKPLKAK--GRSKGILNCQKSTGNSKSPKGLKTPK-----TKMKQMTLDMAK- 445

1779 KSEVAKTKESGLVKRVVEAEKKVTEAR--PKLDAERAKEAVVYVLAQIAMNKKMILTSASV 1837

446 ---GTOQMTRAP---RNSGGTPTRTSKPHKLPALHIAIYYKENKREDKRSALSCVI 499

Db 1838 AILGAGLVASPTVVRAEAPVASQSKAEKD-----YDTAKRDAENAKKAL----- 1883

QY 500 SKTARLLSSEDRARLPPELRSLVQKRYELLEHHKKRWASM-SEEQKRY----- 546

Db 1884 -----BEAKRAQEKYADYQRIEKAETHASLEQOEANKDYQLLKYLDGRN 1933

QY 547 -----LKKBEELKKLKEK----- 561

Db 1934 LNSGVLKKEEAEKDKBEKQAGLMNKKMILTSASVAILGAGLVTSQPTVLVRAESP 1993

QY 562 -----AKERREKEMLERLEKQKRY 580

Db 1994 VASQSKAEKDYDAKDAENAKKALEBAKRAQAEKYADYQRIEKAETHASLEQOEANKDY 2053

QY 581 EDQE-----LTGKVLPAFRIVDTPTEGLPNTLFGDVAMVVFPLSCYSGLLIPDAQYPTA 634

Db 2054 KDYQLKKYLDGRNLSNSVL--KKEMEEAEKKDKKQABFNKIRREIVVNPQELEMA 2111

QY 635 VSLMEALSADKGGF-----LYLNRVLVILQ--TLLODEIAEDYGEIGM 676

Db 2112 RRSKSEVVRKAKSSGLVKRVVEAEKKVTEARQKLDABERAEKVVLPTRVENEV-----HKLXQ 2167

QY 677 KLSEIPLTLHSVSELVRLCLRRSDVQESSESGSDTDNDKDSAAFEDNEVQDEPLEKLTSE 736

Db 2168 KXNKKMILTSASVAIL-----GAGLV-----SQPTFVRAESPQVVEKSSLEK 2213

QY 737 FFELTSEBKQILTALCHRIILMTYSVQDHMETROQMSAELWKLAVLKEENDKGRASKQ 796

Db 2214 KYE-----EAKKADTA-----KKDYETAKKKAEDAQK-----KYEDDQKRTBEK 2253

QY 797 KRKMEAKN-----KENGKVENGLGK-----TDRKKRIVKPEQVD---TEAE 836

Db 2254 ARKEAASQKILIDVALVVQNAVYKEYREVQNRSKYSADYQKLTVEVDSKIEKARKSQ 2313

QY 837 DM-----ISAV-----KSRLL-----AIQAKKER 856

Db 2314 DLQNNFNEVRAVADPTCVGDXDXEMNKKMILTSASVAILGAGXVTSQPTXVRAEAP 2373

QY 857 EIQR-----EMVKLE-----RQAEERIRKHKAABEKAFOEG 890

Db 2374 QVVEKSSLEKKEEAKYDAKKDYDEAKKAAEAKKYBEDQKTEBEKAKAASBE 2433

QY 891 TAKAKLVMRTP---IGTDNRNHRVWLPFSDVPGFLIEKGWVHSDIDYRFNHHCDDHTVS 947

Db 2434 IAKATEEVQKAVLDYITAIRNHDSGKTSAE-----BAENKA 2470

QY 948 GDEYCPRSKK-----ANLGNASMTQHGTTATEVAVET----- 981

Db 2471 KERDYCCAGKFPDPIQTPFVASLTQMLMKNKKMILTSASVAILGAGLVASSPTVVRAE 2530

QY 982 -----TTPKQGNLWFLCD SOKELDELNCLHPOGIRISQLEKRYQ 1026

Db 2531 EAPVASQSKAEKDYDTAKRDAEN-----AKKALEBAKRAQAEKYADYQRIEKAETHAS 2584

QY 1027 DIIHSILHARXENGLKS-CDG-----NOELLNFLRSLDIEVA 1063

Db 2585 ASLEQOEANKDYQLKYYLDGRNLSNSVLKMEEAEEKDKENQAEKFNKIRREIV----- 2641

QY 1064 TRLQKGGGLGYVEETSEFEAR-----VISLEKLDGFCGVIALQASVIKFLOGFMAPK 1116

Db 2642 -----VNPQOELEMANMKMILTSASVAILGAGFVASQPTVVR-----AEESP 2687

QY 1117 QKRKLQSEDSAKTEVEDEBKQVVEAKVA-----SALEKW-----KTAIREAQTF 1162

Db 2688 ASQSKAEKDYDAKDAENAKKAVEDAQKALDDAKAAQKYDEDQKTEEKAALKAASE 2747

QY 1163 SEMVLLGLWLDACIKWDMASNAENARCKVCPKGGEDDKLILCDECNK---AHLFLCLRPAL 1218

Db 2748 EMDKAAVAQVQAYLAYQQAOTDKA-----AKDAADKMI--DEAKKREBAKTKFTVTRAM 2799

QY 1219 YEVPDGEQCPACQAPATARRNSRGRNYTEESASEDEDEDEDEDEDEDEDEDEDEDEYEV- 1277

Db 2800 V-VPPE-----QLAETKKKSEAKQKAPELTKLEBAKAEKKAATEAKQKVDAM 2852


```
; CURRENT FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CurPatSeqFormatter Version 0.9
; SEQ ID NO 1084
; LENGTH: 1972
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: misc feature
; FEATURE:
; LOCATION: (0)_(0)
; OTHER INFORMATION: Polypeptide Accession Number P35749
US-09-538-092-1084
```

```
Query Match 3.7%; Score 292.5; DB 4; Length 1972;
Best Local Similarity 19.9%; Pred. No. 7.4e-12;
Matches 275; Conservative 214; Mismatches 507; Indels 383; Gaps 58;

QY 75 LKKEFPWYKYL-VLEMVHNTASLEKLVDPFWLEIMTKY-----AVGEBCD 121
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
801 LARKAFKQQQLTAMKVIQRNCAAYLKLNNQWNRLLFTKVKPLQVTRQBEEMQAKDE 860
QY 122 FEVGKEMLVKIVKIHLEKVDDEATEKKS-----DCACDPSDDKENSQIAQDHQK 175
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
861 LQTKTEQOQAB-NELKELEQHSOLTEKNILQLOQAETELYABAEEMRVLAACKQE 919
QY 176 KETVKEDEGRRESINDARRPRKLPSTSLKGERKWAPPKPLPHKYDVKLQNEDKIISN 235
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
920 LEEILHEARLEEDRQOQ-----LQAEKK-----KVAQOQLDEEQLER----- 961
QY 236 VPADSLIRTPNKKEIVRYFIRHNALRAGTGENAPWVVEDLVKYSLSPKFS---DPL 292
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
962 -----EEMAROKLQ---LEKVTAEAKIKKLEDEILVMDQNNKLSKERKLL 1004
QY 293 LDPKYMTLNPSTKRNKTSPPRPSKSKTNSLSPLNPKLWCHVHLKXLSG----- 348
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1005 EERISDLTNTLAEBEKA-----KNLTCLKNKHESMISEL-----EVLKKEEKSROEL 1053
QY 349 SPLKVNKSNKSPPEHLEEMKMMKSPNKLHTNFHPKPGPAKPGKHSKPLKAKGRS 408
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1054 EKLRKLEGDADFHEQIADLQAAELKM---QLAKKEELQALARLDDEIAQKNA 1109
QY 409 -KGI-----LNGKSTGNSKSPKK-----GLKTPKTKMKTLLDMAGTKQK 449
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1110 LKKIRELEGHISDLQEDLDSERAARNKAKQKRDILGEELEALKTELED--TLD-STATQQ 1166
QY 450 MTRAPNSGGT-----PRTSKPKHLPALHLLIAYYKENKDEKRSALSVCISKT 502
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1167 ELRAREQEVTVLKALDESTRSHEAQVQEMRQKAQAVEELTEQLQPKRA-KANLDKN 1225
QY 503 ALLSSSEDRARLPPEELRSVQKRYELLEHLKKEKRWASMSBQRKEYLKKREELKXKLEKA 562
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1226 KOTLEKEN-ADLAGELRVLCQAKQE-VEHKK-----KLEAQVQELQSKCSD-- 1270
QY 563 KERREKEMLERLEKQRYDBQLTGKLPALFRLVPTPEGLPNTLFGDVAWVVFELSCVSG 622
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1271 GERARAEINDVKHLQN-EVESVTG-----MLNEAECKAIKAKDVASLSQLQDTQE 1322
QY 623 LLLPQAQYIPITAVSLMEALSADKGGFLYNRVILVILLQTLQDETAED-----YG 672
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1323 LLQETETROQLNVSTKLRLQEEERNSS-----LQDQDEMEAKQNLERHIS 1367
QY 673 ELGKMLSEIPLTLHVSVELVR-LCLRRSDVQSESGSDTDDNDKDSAAPE-----DNEVD 726
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1368 TINIQLSDSKKLQDPFASVTEALLEGKFPQKEIENLTQOYEKAAAYDKLSEKTNRLQ 1427
QY 727 EF-----LEKLETSEFFELTSEKQLQILTALCHRLMTYVSQDMETMQ 770
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1428 ELDDLVDLDNQRLVSNLEK-KQRKFOQLAABEK-----NISSKYA-----DERD 1472
```

```
QY 771 QMSAEL-WKERLAV-----LKEENDKKRAEKQKQKEMEAKNKENGKVENGLGKTDR 820
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1473 RAEAEAREKETKALSLARALEALEKEELRTNKMLEAEMEDLVSSKDDVGVKNVHELEK 1532
QY 821 KKRIVKFPQVD-----TEADMISAVKSRRL-----LAIQAKKEREIOBR-----BM 863
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1533 SKRAL--ETOMEEMKTQLEEELEDELQATEDAKRLLEVNNQALKGOFERDLOARDEQNEEK 1590
QY 864 KVKLERQ-----AEERIRKHKAARAKAQEGIAKAKLVWRRTPPI-GTDENHNRYWLF 915
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1591 RRQLQRQLHEYETELEDERKORALAAAKKLEGLDKOLEQADSAINGRE----- 1641
QY 916 SDEVPLFIEKGWVHDSIDYRFNHHCKDHTVSGDDYC---PRSKKANLGKNASMNTQHG 972
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1642 -EAKQLRKLQAKWD-----FORELEDARASRDEIFATAKENEKAKSLEADLMQLQED 1695
QY 973 TAT-----EVAVETTTTPKOGNLMWFLCDSQKELDELNCLHPQGISOLKE 1019
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1696 LAAERARKQADLEKEELAEELASSLSGRNA-LQDEKRRLE-----ARIAQLEE 1743
QY 1020 RLEKRYQDIHIIHLARPNGLKSCDQNOELLNRLSLDLIEVATRLQKGLGYVEETSE 1079
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1744 ELEEB-----QGNMEAM-----SDRVKAKATO----- 1764
QY 1080 FEARVISLEKLDGFCGVIALQASVIKFLQGFMAPKOKRRKLOQSEDSAKTEEVDEEKKM 1139
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1765 -QAEQLSNE-----LATERSTAQK-----NESARQQLERQNKELSKLHEMGA 1807
QY 1140 VE-----EAKVASALEKWKTAIREAQTFSRMHVLLGLMDACIKWDSAEANARKV 1189
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1808 VKSKFKSTIAALEAKIAQLEEQVEQAREKQAATK----- 1842
QY 1190 CPKGGEDKLLILDCBNKAFHLCPALPYEVPDGEWQCPACQATARRNSGRNYTE--- 1247
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1843 -SLKOKKK-----LKEILLQVEDERKMAEQYKEAQEKGNARVKQLKQQL 1886
QY 1248 ESASDSDDDSDEEBEEEEEBEEDYEAGLRLRPRKT--IRGKHSVIPPAARSGR 1304
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1887 BEAEESORINANRRKLQRELDDEATESNEAMGRENVALKSLRRNGNETSFVPSRRSGGR 1945

RESULT 31
US-09-949-016-7111
; Sequence 7111, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7111
; LENGTH: 1984
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-7111

Query Match 3.7%; Score 292.5; DB 4; Length 1984;
Best Local Similarity 19.9%; Pred. No. 7.5e-12;
Matches 275; Conservative 214; Mismatches 507; Indels 383; Gaps 58;

QY 75 LKKEFPWYKYL-VLEMVHNTASLEKLVDPFWLEIMTKY-----AVGEBCD 121
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
813 LARKAFKQQQLTAMKVIQRNCAAYLKLNNQWNRLLFTKVKPLQVTRQBEEMQAKDE 872
```


771 QMSAEL-WKERLAV-----LKEENDKRAEKQKREMEARKENGKVENGLGKTR 820
1485 RAEAEAREKETKALSARALEEAELEAKELERTNKMLKAEMEDLVSSKDDVGNVHELEK 1544
821 KKRIVKFEQVD-----TEADMI SAVKSRRL-----LAIQAKKERIOER-----EM 863
1545 SKRAL--ETOMEEMKTQLEBEDELQATADAKLRLEVNQALQGQFERDQARDEQNEEK 1602
864 KVKLERQ-----ABEERIRKHAARAFOGIAKAKLVMRTP1-GTDRNHNRYWLF 915
1603 RQLOQLQHEYTELEDERKQALAAAKKLEGLDKDLELQADSAIKRE-----1653
916 SDEVPLGFTIEKGWVHDSIDYRFNHKCHDHTVSGDEYIC---PRSKANLGKNASMNTQHG 972
1654 -BAIKQLRKLQAMKD-----FORELEDARASRDEIFATAKENEKAKSLEADLMQLQED 1707
973 TAT-----EVAVETTPKQGNLWFLCDSQKELDELNLCHLPHQIGRESQKXE 1019
1708 LAAAEARKQADLEKEELABELASSGRNA--LQDEKRRLE-----ARIAQLEE 1755
1020 RLEKRYQDIHSHLARKPNLGLKSCDGNQELNPLRSDLIEVATRLQKGLGYVEETSE 1079
1756 ELEEE-----QGNWEAM-----SRVRKATO-----1776
1080 FEARVISLEKLDGFCVIALQASVIKFLQGFMAPKQKRRKIQSDEDSAKTBEVDDEKQM 1139
1777 -QAEQLSNE-----LATERSTAQK-----NESARQQLERQNKELRSKHEMEGA 1819
1140 VE-----BAKVASALEKWKTAIRQAQTFSRMHVLLGMLDACIKWDMGAENARCKV 1189
1820 VKSFKSTIAALEAKIAQLEEQVEQAEKQAATK-----1854
1190 CPKKGDDKLLCDECNKAFHLCLRLPALYVDPDGEWQCPACQAPATARENSRGNYTE-- 1247
1855 -SLKQDKK-----LKEILLQVEDERKMAEQYKEQAEKGNARVQKLRQL 1898
1248 ESASEDSEDESEDESEDESEDESEDESEDESEDESEDESEDESEDESEDESEDESEDE 1304
1899 EBAEESQRINARRKLRQELDEATESNEAMGREVNALSKLRGRNETSFVPSRRSGOR 1957

RESULT 34
US-08-875-435B-4
; Sequence 4, Application US/08875435B
; Patent No. 6593304
; GENERAL INFORMATION:
; APPLICANT: Haegawa, Kazuhide
; APPLICANT: Arakawa, Emi
; APPLICANT: Oda, Shoji
; APPLICANT: Matsuda, Yuzuru
; APPLICANT: Takahashi, Katsuhito
; APPLICANT: Sugahara, Michihiro
; APPLICANT: Ishiyama, Haruo
; TITLE OF INVENTION: RECOMBINANT DNA COMPRISING DNA CODING
; TITLE OF INVENTION: FOR MYOSIN HEAVY CHAIN SM1 ISOFORM PROTEIN INSERTED INTO
; TITLE OF INVENTION: VECTOR DNA, MICROORGANISM CARRYING THE RECOMBINANT DNA, AND
; TITLE OF INVENTION: AN AGENT FOR TREATMENT OF ARTERIOSCLEROSIS COMPRISING THE
; TITLE OF INVENTION: RECOMBINANT DNA
; FILE REFERENCE: 07898-013001
; CURRENT APPLICATION NUMBER: US/08/875,435B
; CURRENT FILING DATE: 1997-07-25
; PRIOR APPLICATION NUMBER: PCT/JP96/00134
; PRIOR FILING DATE: 1996-01-25
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 1972
; TYPE: PRT
; ORGANISM: Oryctolagus cuniculus
US-08-875-435B-4

Query Match 3.5%; Score 281.5; DB 4; Length 1972;

Best Local Similarity 19.2%; Pred. No. 4.6e-11;
Matches 266; Conservative 219; Mismatches 503; Indels 399; Gaps 60;
QY 75 LKKEBPPAWVEKL-VLEVMYHNTASLEKLVDVTAWLEIMTK-----YAVGECCDFEVGKEK 128
DB 801 LARKAPAKQOQQTANKVIQORCAAYLKURNQWRLFTKVRPLLOVTTQEEBQAKADE 860
QY 129 MLKVK-----IVKIHPLEKVDDEATEKKS-----DGACDSPSSSKENSQIAQDHQKK 176
DB 861 LQKIKERQQAESSELQELQOKHTQOLSEENKLLQEQLAETELVAAEEMVRVLAQKQEL 920
QY 177 ETVVKEDEGRRSINDRARRSPKPTSLKGBERKWPAPKPLPHKYDVKLQNEDKKIISNV 236
DB 921 EBIHEMEARLEEEEDRGQ-----LOAERK-----KMAQQMLDLEEQLEE-----961
QY 237 PADSLIRTERPPNKETVRYFIRHNALRAGTGNAPWVVEDELVKVYSLSKFS-----DFLL 293
DB 962 -----EENAROKLQ-----LEKVTAAKIKLEDDILVMDQNNKLSKERKLE 1005
QY 294 DPYKMTLNPSTKRNKTGSPDRPKSKSTDNSSLSPLNPKLHVHLKSLSG-----S 349
DB 1006 ERISDLTTNLAEEEKA-----KNLTKLNKHESMISEL-----EVLKKEEKSQELE 1054
QY 350 PLKVNNSKNSKSPPEHLEEMMOGPNKLUHFNPHIPKGGPPAKKP-----GKHGXPL 402
DB 1055 KLKRXMDGEASDLHEQIADLQAIABLKML-----AKKEEELQAALARLEDETS 1104
QY 403 KAKGRSKGT-----LNGOKSTGNSKSPK-----GLKTPKTKMKQMTLLDMA 444
DB 1105 QKNALIKTIRELGHISDLQEDLDSERAARNKAQKQRLGEBLEAKTLEED--TLD-T 1161
QY 445 KGTQKMTAPRNSGGT-----PRTSS-----KPHGLPPAALHUIAYYKKNKOR 488
DB 1162 TATQELRAKREQEVTVLKALDEETRSHEAQVQEMRQKHTQVVEBELTQEQFKEAKAN 1221
QY 489 EDRSALS CVISKTARLJSSDRARLPPEELSLVQKRYELLBHKHKKWASMESEQRKEYLK 548
DB 1222 LD-----KTKQTTLEKEN-ADLAGELRVLQAKQES-VEHKKK-----KLE 1258
QY 549 KXREELKKLKEKAKERREKEMLEKOKRYVEDQELTGKNLPAFLVDTPEGLPNTLFG 608
DB 1259 VOLQELQSKCD--GERARAEINDKVHKLQN-EVESVTG-----MUSEAEKATKLAK 1308
QY 609 DVAMVVEFLSCYSGLLLPDAQYPIITAVSLMEALSADKGGFLVNLRVLLTQLLQDETA 668
DB 1309 EVASLGSQLODTQELLQEBETROKLVNSTKLRLQLEDERNS-----LQELQDEME 1357
QY 669 E-----DYGELGMKLSEIPLTLHVSSELVRLCLRRSDVQESSESDTDNDKD 715
DB 1358 AKQNLERHISTLNIQLSDSKKKLQDFASTVESLEE-----GKKRFQKEISLTQOYEEK 1411
QY 716 SAAFE-----DNEVODEF-----LEKLETSEFFELTSEKQLILTALCH 754
DB 1412 AAAYDKLEKTKRVLQOELDDLVLDNQRLVSNLEK-KOKKFDQLLAEBK-----1461
QY 755 RILMTYSVQDHMETRQMSAEL-WKERLAV-----LKEENDKRAEKQKREMEARK 804
DB 1462 NISSKYA-----DERDRAEAEAREKETKALSARALEEAELEAKELERTNKMLKAEMEDL 1516
QY 805 NKENGKVENGLGKTRKRIKRVKFEQVD-----TEADMI SAVKSRRL-----LAIQA 852
DB 1517 VSSKDDVGNVHELEKSKRAL--ETQMEEMKTQLEBEDELQATADAKLRLEVNQALQV 1574
QY 853 KKERIOER-----EMKVKLREQ-----ABEERIRKHAARAFOGIAKAKLVMR 900
DB 1575 QFERDQARDEQNEEKRRQLQRLQHEYTELEDERKQALAAAKKLEGLDKDLELQAD 1634
QY 901 TPI-GTDRNHNRYWLFPSDEVPLGFTIEKGWVHDSIDYRFNHKCHDHTVSGDEYIC---PRS 956
DB 1635 SAIKGR-----FAIKQLLKLQAMKD-----FORELEDARASRDEIFATAKENE 1679
QY 957 KXANLGKNASMNTQHGTTAT-----EVAVETTPKQGNLWFLCDSQKELDEL 1003

Db 1680 KKAQSLADLMQLQEDLAAARQADLEKEELAEELASSLSGRNA--LQDEKRRLE-- 1735
Qy 1004 LNCILHPOGIRESQLERLEKRYQDIHSHILARKENLGLKSCDGNQELLNPLRSLLIEVA 1063
Db 1736 -----ARIAQLEEELEEE-----QGNMEAM-----SDRVKRA 1762
Qy 1064 TRLQKGLGYVBETSEFEARVISLEKLDKDFGCVIALQASVKIKFLQGFMAPKQRRKLIQ 1123
Db 1763 TO-----QAQLSNELAT-----RSTAQ 1781
Qy 1124 SEDSAKTEVDEBKQVKEAKVASALEKWKTAIRE--AQTFSRMHVLLGMLDACIKWDM 1181
Db 1782 KNEAR-QQLERONK-----ELSKLQEMEGAVKSKFKSTIAALEAKIALEEQVEQEAR 1835
Qy 1182 AENACKVCPPKGGDDKLIILDECNKAFHLFCLRLPALYVPPGEGWQCACQATARRNSR 1241
Db 1836 EKQAAKAL--KORDKK-----LKEMLLQVEDERKVAEQYKEQAEKGNAK 1878
Qy 1242 GRNYTE--ESASEDESDDEEBEEREEEDYEVAGLRRLRPRKT--IRGKHSVIPP 1297
Db 1879 VKQLKRLQLEAEESQRINANRRKLQRELDDEATESNEAMGREVNALKSKLRGNETSFPV 1938
Qy 1298 AARSRR 1304
Db 1939 TRRSGR 1945

RESULT 35
US-09-418-710-68
; Sequence 68, Application US/09418710
; Patent No. 6596482
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042001
; CURRENT APPLICATION NUMBER: US/09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 68
; LENGTH: 65
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-418-710-68

Query Match 3.5%; Score 281; DB 4; Length 65;
Best Local Similarity 92.3%; Pred. No. 3.8e-13;
Matches 60; Conservative 1; Mismatches 0; Indels 4; Gaps 1;

Qy 630 YPITAVSLMEALSADKGGFLYNRLVLI-----LLQTLQDEIAEDYGEIGMKLSIPITL 685
Db 1 YPITAVSLMEALSADKGGFLYNRLVLI-----LLQTLQDEIAEDYGEIGMKLSIPITL 60
Qy 686 HSVSE 690
Db 61 HSVSE 65

RESULT 36
US-09-839-479-67
; Sequence 67, Application US/09839479
; Patent No. 672722
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042002
; CURRENT APPLICATION NUMBER: US/09/839,479

; CURRENT FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 67
; LENGTH: 65
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-839-479-67

Query Match 3.5%; Score 281; DB 4; Length 65;
Best Local Similarity 92.3%; Pred. No. 3.8e-13;
Matches 60; Conservative 1; Mismatches 0; Indels 4; Gaps 1;

Qy 630 YPITAVSLMEALSADKGGFLYNRLVLI-----LLQTLQDEIAEDYGEIGMKLSIPITL 685
Db 1 YPITAVSLMEALSADKGGFLYNRLVLI-----LLQTLQDEIAEDYGEIGMKLSIPITL 60
Qy 686 HSVSE 690
Db 61 HSVSE 65

RESULT 37
US-09-595-684B-31
; Sequence 31, Application US/09595684B
; Patent No. 6544766
; GENERAL INFORMATION:
; APPLICANT: Beraud, Christophe
; APPLICANT: Ohashi, Cara
; APPLICANT: Sakowicz, Roman
; APPLICANT: Vaisberg, Eugeni
; APPLICANT: Wood, Kenneth
; APPLICANT: Yu, Ming
; TITLE OF INVENTION: Human kinesins and methods of producing
; FILE REFERENCE: cytop036
; CURRENT APPLICATION NUMBER: US/09/595,684B
; CURRENT FILING DATE: 2002-06-24
; PRIOR APPLICATION NUMBER: 09/295,612
; PRIOR FILING DATE: 2000-04-20
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 31
; LENGTH: 2662
; TYPE: PRT
; ORGANISM: Human
US-09-595-684B-31

Query Match 3.5%; Score 281; DB 4; Length 2662;
Best Local Similarity 19.3%; Pred. No. 7.6e-11;
Matches 337; Conservative 302; Mismatches 598; Indels 508; Gaps 82;

Qy 39 EARLERYSERIWTCKSTGSSQLTHKEAWEEEOEVAELKEEPPAWYKVLVLEMVHNTAS 98
Db 788 ESRVQGLLEEIGTKTDDLATTSQSNYKSTDOEFQNTLHMDFEQKY-KWVLE---ENERM 843
Qy 99 LEKLVDTAWLEIMTKYAVGECDFEVGKEKMLKVKI-VKIHPLEKVDDEBATEKSDGACD 157
Db 844 NQEIWN-----LSKEA--QKFDSSLG---ALKTSLYKTOBLOKXTREVQERLNE---- 888
Qy 158 SPSSDKENSSQIAQDHQKKEVTVVKDEGRRESINDRRSPKLPY-SLKKGKRWAPPK 216
Db 889 -----MEQLKEQLENRDSPLQTVREKTLITEKLOQTLVEVKTLTQEKDQLQLOES 940
Qy 217 FLPHKYDKVLQKQEDKIISNVPADSLIR-----TERPPNKIVRYFIRHNLRACT 266

Qy	158	SP	SD	KENS	QIAQ	HQ	QK	ET	VV	KE	DE	GR	ES	IN	DR	RR	SP	KL	PT	--	SL	KK	GR	KW	AP	PK	216	
Db	889	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	940	
Qy	217	FL	PH	KY	DV	KL	ON	ED	KI	I	S	N	P	A	D	S	L	I	R	----	----	----	----	----	----	----	265	
Db	941	LQ	IR	DQ	LK	S	D	I	H	T	V	N	N	I	D	T	O	B	L	R	N	A	L	E	S	K	1000	
Qy	266	TG	NA	PW	VE	DE	L	V	K	Y	S	L	P	S	K	S	D	F	L	L	D	P	V	K	Y	T	325	
Db	1001	TGE	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	1053	
Qy	326	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	382		
Db	1054	EL	Q	M	L	S	V	A	E	R	E	Q	L	T	D	L	K	E	N	I	----	----	----	----	----	----	----	1108
Qy	383	HI	P	K	G	P	A	K	P	G	H	S	D	K	P	L	K	A	G	R	S	K	I	L	N	G	432	
Db	1109	-I	K	E	G	E	L	S	R	T	C	O	R	A	E	E	K	E	S	O	O	L	N	V	O	E	M	1166
Qy	433	T	K	M	O	M	T	L	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	462	
Db	1167	L	K	O	K	E	L	T	E	M	E	T	E	R	L	E	A	O	K	N	E	N	E	V	E	S	I	1226
Qy	463	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	514	
Db	1227	E	A	T	G	L	O	T	K	E	L	I	A	H	I	L	----	----	----	----	----	----	----	----	----	----	1278	
Qy	515	P	E	L	R	S	I	V	O	K	R	Y	E	L	L	E	H	K	K	R	W	A	S	N	S	----	----	551
Db	1279	Q	E	E	I	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	1337	
Qy	552	E	L	K	K	L	K	E	K	A	R	R	E	----	----	----	----	----	----	----	----	----	----	----	----	----	592	
Db	1338	Q	E	S	E	E	I	K	S	I	T	K	E	R	D	N	L	T	I	K	E	A	L	E	V	K	H	1394
Qy	593	F	R	L	V	D	T	P	E	G	I	P	N	T	L	F	G	D	V	A	M	V	E	L	S	C	S	642
Db	1395	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	1441	
Qy	643	A	D	K	G	F	L	N	R	V	L	I	L	O	T	L	O	----	----	----	----	----	----	----	----	----	----	698
Db	1442	K	E	D	D	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	1489	
Qy	699	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	723	
Db	1490	E	E	T	I	N	E	L	R	N	L	S	E	K	E	T	E	I	S	T	I	Q	K	L	E	A	I	1549
Qy	724	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	769	
Db	1550	L	K	O	F	K	E	H	R	K	A	D	S	A	L	O	I	E	----	----	----	----	----	----	----	----	----	1605
Qy	770	Q	M	S	A																							

Qy	1034	L	ARKPNIGL	KSCDGN	QELN	FLRSD	LIE	VATRL	-	QKGG	LYV	BET	S	E	F	E	A	R	V	I	S	L	E	K	L	D	1092																													
		:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:																														
Db	1863	L	S	K	--	L	E	I	E	N	L	N	A	Q	E	---	L	H	E	N	E	E	M	K	S	V	M	K	E	R	D	N	L	R	V	E	T	L	K	L	E	---	D	Q	L	K	E	1910								
		:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:							
Qy	1093	F	G	E	C	V	I	A	L	O	A	S	V	I	K	F	L	O	G	F	M	A	P	K	O	X	R	---	K	L	O	S	E	S	A	K	T	E	V	D	E	K	M	V	E	A	K	A	S	A	L	1149				
		:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:							
Db	1911	S	L	O	E	T	K	A	R	D	L	E	I	O	E	L	K	T	A	R	M	L	S	K	E	H	K	E	T	V	D	K	L	E	R	K	I	S	E	K	T	I	O	I	S	I	Q	I	O	D	L	D	K	S	---	1965
		:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:							
Qy	1150	E	K	W	K	T	A	I	R	E	A	O	T	S	R	M	H	V	L	L	G	M	D	A	C	I	K	W	D	M	S	A	E	N	A	R	---	C	K	V	P	K	G	E	D	D	L	I	L	D	E	C	N	1206		
		:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:							
Db	1966	D	E	L	O	K	I	O	E	L	O	K	-	K	E	L	L	R	---	V	K	D	V	N	S	H	K	I	N	E	M	E	O	L	K	K	O	F	E	P	N	Y	L	C	K	E	M	D	N	1208						
		:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:							
Qy	1207	K	A	F	H	L	C	L	R	P	A	L	E	V	P	D	G	E	W	C	P	A	C	O	P	A	T	A	R	R	N	S	R	G	N	T	E	S	A	S	E	D	S	E	D	E	S	E	E	E	E	1266				
		:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:							
Db	2019	---	F	O	L	T	K	L	H	S	L	E	-	E	I	R	I	V	A	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---			
Qy	1267	E	B	E	E	E	E	D	Y	V	A	G	L	R	P	R	K	T	I	G	K	S	V	I	P	P	A	A	R	S	G	R	R	P	O	K	---	P	H	S	T	R	R	S	O	P	K	A	P	1320						
		:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:							
Db	2047	E	S	L	K	M	E	R	D	Q	I	A	T	L	R	E	M	I	A	R	-																																			

QY 690 E 690
Db 59 E 59

RESULT 40
US-09-839-479-66
; Sequence 66, Application US/09839479
; Patent No. 672722
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042002
; CURRENT APPLICATION NUMBER: US/09/839,479
; PRIOR FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 66
; LENGTH: 59
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-839-479-66

Query Match 3.5%; Score 277; DB 4; Length 59;
Best Local Similarity 96.7%; Pred. No. 6.5e-13;
Matches 59; Conservative 0; Mismatches 0; Indels 2; Gaps 1;

QY 630 YPTAVSLMEALSADKGGFLYLNRLVILLQTLQDEIAEDYGEIKNLSEIPLTLHVS 689
Db 1 YPTAVSLMEALSADKGGFLYLNRLVILLQTLQDEIAEDYGEIKNLSEIPLTLHVS 58

QY 690 E 690
Db 59 E 59

RESULT 41
US-09-538-092-1246
; Sequence 1246, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
; APPLICANT: Mansfield, Traci A.
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538,092
; CURRENT FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CurPatSeqFormat Version 0.9
; SEQ ID NO 1246
; LENGTH: 3924
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: Polypeptide Accession Number Q01484
; NAME/KEY: misc_feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: Polypeptide Accession Number Q01485
US-09-538-092-1246

Query Match 3.4%; Score 274.5; DB 4; Length 3924;
Best Local Similarity 19.5%; Pred. No. 3.9e-10;
Matches 348; Conservative 243; Mismatches 673; Indels 519; Gaps 81;

QY 19 EPPFTIPTQEAFTREYERARLYSE--RIWTC-KSTGSSQLTHKEAMESEQVAVEL 75
Db 1510 EEP-----GEPFEIVERVKEDLEKVNELRSLGTCTRDSESSVQSSRSRGLVEEWVIV 1562

QY 76 LKEEPPAWYKLVLEWVHNTASLEKLVDTAWLEIMTKYAVGEECFEYGEKMKLVKIV 135
Db 1563 SDEEIEEARQKAPLE-----ITEY-----PC-VEVRIDKEIKGKVE 1597

QY 136 K-----IHPLEKVDEEATEKKSDGACDS-----PSSDKE-----NSSQIA 170
Db 1598 KUSTGLVNYLTDLNTCVPLPKSQLQVQDKAGKACEALAVGRSSEKGGKIDIPDETQST 1657

QY 171 QHQKQKTVVKDEGRRESINDRARRSPKLPSTLKKGKRWAPPKPLPHKYDVKLQNEED 230
Db 1658 QKHKPSLGIKKPV--RRKLKEKQKQKEGLQASAEKELK-----KGSSEE 1702

QY 231 KI-----ISNVPADSLIRTEPPN--KEIVRYFIRHNALRAGTGENAPWVVEDEL 278
Db 1703 SLGEDPGLAPEPLPTVKATSPLEIETPIGSIKDKVKALQKR-----VEDEQ 1748

QY 279 VKKYSLPSPKSFDFLLDPYK-----YMTLNPSTK--RKNTGSPDRKPSKSKTSDNSLSS- 330
Db 1749 KGRSKLPPIRVKGKEDVPKKTTHRPHPAASPPLKSERHAPGSPSPKTERHSTLSSAKTER 1808

QY 331 --PLNPKLWCHVHLKXLSGSLPKVKNKSNKSPSEHLEEMMMOMSPNKLHNTFHP--- 385
Db 1809 HPPVSPSSKTEKHSPVSPSAKTERHSPASSSSSKTEKH-----SPVSPS-TKTERHSPVSS 1862

QY 386 ---KKGPPAKKQKHSKDK--PLKAKGRSKGILNGQKSTGNSKSPKGLTKPTKMKQMTL 440
Db 1863 TKTERHPPVSPSGK-TDKRPPVSPSGRTE-----KHPP-----1894

QY 441 LDMAGTKQMTAPRNSGGTTPR-----TSSKPKHLLPAAALHLIAYKENKREDKRSAL 495
Db 1895 VSPGRTEKELPVSP--SGRTDKHQPVSTAGTEKHLVPSP-----SGTEKQPPV 1942

QY 496 SCVISTARLLSSEDRARLPBELRLVQRYELLEHKRWASMBEQREKYLKKKEELK 555
Db 1943 S-PTSSTERI-----EETMSV-----RELM--KAFQSGQDPSKHKHTGLFEHKSQK 1985

QY 556 KKLKEKAKERREKE---MLERLEKOKRYEDQELTGKLPAPFLVDPTEGLPNTLFG---- 608
Db 1986 KQPEKGRVVRVEKEGPILTOREAOKTENQTIKRGORLPVTGTAEKRGVRVSSIGVYKKE 2045

QY 609 DVAMVVEFLSCYSGLLLPDAQYPIITAVSLMEALSADKGGFLYLNRLVILLQTLQDEIA 668
Db 2046 DAAGGKERVLSHK--IPE---PVQSVPEESHRESE-----VPEKMA 2083

QY 669 EDYBELGKMLSEIPLTLHVSSELVRLCLRRSD-----VOEESGSDTD-----DNKDSA 717
Db 2084 DEQGDMDLIQSPDRKTSTDFSEVIKQELBDNDKYQQFRLSEETEKALHLDQVLTSPPFNT 2143

QY 718 APEDNEVDQEFLEKLETSEFFELTSEEKQ---ILTALCHRLIMTYSVQDHMETROMSA 774
Db 2144 TFPDYMKDEFLPALSLQSGALDGSSESLKNEGVAGSPCGSL-----MEGTQIIS 2194

QY 775 E--LWKEKRLAVLKEENDKKRA-EKOKRKEAMEAKNKGKENGVLGKTRDKKRVKPEQV 831
Db 2195 EESYKHGELAETPETSPLSPSPKSEOTGETESTYTE-----TTTEISEKEHP-- 2247

QY 832 DTEAEDMISAVKRRLLAIQAKKEREIQEREMKVKLERQAEEREIRKKHAAKAFQEGE 891
Db 2248 --TTKDIITGGSBERGATVTE-----DSETSTESFQKEATLGS PKDTSPKRDDCT 2295

QY 892 AKAKLVM-RRTPIGTDNRNRYNLWLSDEVPGLFIEKGWVHDSIDYRFNHCHDKHTVSGDE 950
Db 2296 GSCSVALAKETPTGLTEE-----AACDEGQRTFGSS--AHKT---QTDSEAEQSTATSD 2345

```
QY 951 DYCPRSKANLCKNKMNTQHGTFATEVAVETTTBKQGNLWFLCDSQKEL-----DELLNC 1006
Db 2346 -----TKALPLPEASVKTTGTESKPGQVIRSP-QGLELALPSRDSSEVL SAVADSLAV 2398
QY 1007 LH-----PGIRESQLKERLEKRYQDI IHSIHLARKPNLG-----1041
Db 2399 SHKDSLEASPVLEDNSSHKTPSLSPSLK--ESPGRDLSLESSPVEPKKAGIFPSHP 2455
QY 1042 LKSCDGNQELLNPLRDLLEVATRIQKGLGVVEETSEFEARVISLEKLDKDFGECVIALQ 1101
Db 2456 LPAAVAKTELL-----TEVASRSLRLDDPGSAEDD-----SLE-----Q 2491
QY 1102 ASVIKKFLOGFMAP-KQKERKQSDSACT-----BEVDEEK-----1137
Db 2492 TSLMESSGKSPSPDTPPSSESVSEVETPKTTDVSTPKPAVIEHCEAEEDSENGEKKRFT 2551
QY 1138 -----KMVEEAKVASALEK-----WKTATIREAQTFSRMHVLLGLMDACIKWMSAENA 1185
Db 2552 BEEMFQWTKIKMFDELEQAKQKRDYKKEPKQESS-----SSDPDA 2595
QY 1186 RKVC---PK---KGEDDK---LILCDECNKAFHLFCRLPALLYEVPDQ-----E 1225
Db 2596 DCSVDVDEPKHTGSGEDESVPVLTSESRSKVSSESSEPELAQLKKGADSGLLPEPVIR 2655
QY 1226 WCPACQOPATARRNR-----GRNYTESASEDESDESEEEEEEEEEEDY 1275
Db 2656 VOPPGLPSSMDSNSSPEEVQPVVSKQYTFK-MNEDTQEEPGKSEEEKSESHLAEDR 2714
QY 1276 EVAGLRRLPRKTRIRGHSVIPPAARSGR-----RPGKKPHSTRRSOPKAPPVD-----1323
Db 2715 HAVSTEADRSYDKLNRDTPDKICDGHCEAMSPSSSARPYSSGLQSTGDDVDDEQPI 2774
QY 1324 -----DAEVDLVLQTKRS-----1337
Db 2775 YKESLALQTHEKDTGEGBELDVSRAESPOADCPSESFSSSSSLPHCLVSEKGLDEDISA 2834
QY 1338 -SRRQSLQKCEIILHKIVK-----YRFWSW--PFRPVTRDEAEDYDVIT 1381
Db 2835 TSSIQTEVTKTDETFENLPKCPQDSSITQTDRFSMDVPVSLAENDEI--YDPQIT 2892
QY 1382 HPMDFOTVQNKSCGSYSRVQBFLLTDMKQVFTNAEYVNCGRSHVLSMWKTEQCLVLVLIH 1441
Db 2893 SP--YENVPSQSFSSSESKTQTDANHTTSFHSSEVYVITSPVEDVVVASSSGTVLS 2950
QY 1442 KHLPGHYVRRKRKPPDRLADEGDS-----EP-----EAVGQ-----SRDEDR 1482
Db 2951 KESNFEQDQIKMESQLESTLWEMQSDSVSSSEPTMSATTVVGEQISKVITTKTDVSD 3010
QY 1483 SREAEIQEWLQDTSLSYSAKINSKDHNCFMMLVNTQFCMALTDT 1525
Db 3011 S-WSEIRE---DDEAPEARVKEEQKIFGLMWDRSQGTTPTD 3049

RESULT 42
US-08-056-200-94
; Sequence 94, Application US/08056200
; Patent No. 5616500
; GENERAL INFORMATION:
; APPLICANT: Steinert, Peter M.
; APPLICANT: Lee, Seung-Chul
; APPLICANT: Kim, In-Gyu
; APPLICANT: Chung, Soo-Il
; APPLICANT: Park, Sang-Chul
; TITLE OF INVENTION: Trichohyalin and Transglutaminase-3 and
; TITLE OF INVENTION: Methods of Using Same
; NUMBER OF SEQUENCES: 117
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 620 Newport Center Drive, Sixteenth Floor
; CITY: Newport Beach
; STATE: CA
; COUNTRY: U.S.A.
; ZIP: 92660

;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/056,200
; FILING DATE: 30-APR-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Fedrick, Michael F.
; REGISTRATION NUMBER: 36,799
; REFERENCE/DOCKET NUMBER: NIH054.001A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (714) 760-0404
; TELEFAX: (714) 760-9502
; INFORMATION FOR SEQ ID NO: 94:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1898 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-056-200-94

Query Match 3.4%; Score 274; DB 1; Length 1898;
Best Local Similarity 18.4%; Pred. No. 1.5e-10;
Matches 192; Conservative 192; Mismatches 451; Indels 206; Gaps 30;

QY 483 KENDREKRSALSCVISKTARLLSSEDRALPEELRSLVKRYELLEHKKWASMSPEQ 542
Db 288 RERQEEQQQLRL--REQQLRRKQEEERREQERREQERREQERREQERREQERRE 345
QY 543 RKE-VLKKRRELKXK---LKEKAKEREKEML-----ERLEKQRYEDQELTKNLP 593
Db 346 RREQLRQEEERREQERREQERREQERREQERREQERREQERREQERREQERRE 405
QY 594 RLVDTPGLPNTLFGDVAMVVEFLSCYSGLLLPDAQYPITAVSLMEALSADGGFLY 653
Db 406 QQLRREQQLRRE-----QQLRREQQLRREQ-----QLRR 434
QY 654 VLVIILLQTLLODEIAEDYGELGMLKSEIPLTLHSVSELVRLCLRSDVQEESE 713
Db 435 EQQLRREQQLRREQEEERHEQHEQERREQLRQEEERDMLKR---EETERHEQ 491
QY 714 KDSAAFENEVDQEFLEKLETSPEFELTSEELQLITLCHRLMTYSVDHMEIR-Q 772
Db 492 KQQLRQDEEERERRLKLE-----EER-----REQQERREQQL 526
QY 773 SAEWLKERLAVLKEENDKKRAEKQKRAKEMEAQKNGKVENGLGKTDRK----- 827
Db 527 RREQERREQLRQEEERLQQLRSEQLRREQERLEQLKREEEKKLEQERREQRL 586
QY 828 EPQVDEADBMISAVKSRRLLAIAQAKEREIQEREMKVKLEQAAE-----ERIR 883
Db 587 KREQEERDQLLKREERERQQLRQEEERLEQLRKEEVERLEQEEERDERLKRPEE 646
QY 884 EXAFQEGITAKLVMRPTPIGTDNRHNYWLFSEVPGLFIEKGWVHDSIDRFNHCK 943
Db 647 ERRELLKSEQEERHEQLRREQERREQLRKEE-----EERLEQLKKEHEERE 701
QY 944 HTVSGDEDYCPRSK-KANLGKNASMTQGTATEVAVETTTKQGNLWFLCDSQKEL 1002
Db 702 QELAEQEQAERIKSRIPKQWQLESEADARQSKVLEAQAQR-----AEAQEE 756
QY 1003 LLNCLHPQIGRESQL-----KERLEKRYQDI IHSIHLARKPNLGKSCDGNQELL 1058
Db 757 -----KRRSESELQWQEEERAHQEQEEERQDFTWQAEKSEKSERQRL--SAR 807
QY 1059 LIIEVATRIQKGLGVVEETSEFEARVISLEKLDKDFGECVIALQASVIKKFLOG 1118
Db 808 LREQERQLR-----AEERQERQFLPEEKE-----836
QY 1119 RRLQSEDSAKTEEVDEEKKWVEEAKVASALEKWKTAIREAQTFSRMHVLLGLMDACI 1178
```


QY 136 K-----IHPLEKVBDEATEKKSQDACS-----PSSDK-----NSSQIA 170
DB 1425 KDSSTGLVNYLTDLNLCVPLPKBQLQTVQDKACKCEALAVGRSSEKGGKIDPPDETQST 1484
QY 171 ODHOKKQTVVKEDEGRESINDRARRSPKLPSTLKKGERKWAAPPFLPHKDVKLONED 230
DB 1485 QKHOKPSLGKIKFV--RRKLKQKQKEBGLQASAEKALK-----KGSSE 1529
QY 231 KI-----ISNPADSLITERPPN--KEIVRYFIRHNALRAGTGENAPVWVEDEL 278
DB 1530 SLEDGPLAPEPLPTVKATSPLEETPIGSIKDKVKALQK-----VEDSQ 1575
QY 279 VKYISLPKSPFSLDPYK-----YMTLPSTK--RKNTGSPDRKPSKSKTDNSLSS-- 330
DB 1576 KGRSKLPPIRVKGEDVPKKTTHRPHPAASPLKSRHAPGSPSPKTERHSTLSSAKTER 1635
QY 331 --PLNPKLWCHVHLKSLGSPKLVKNKSNKSPKEHLEMMKMWPNKLTNTHFHP---- 385
DB 1636 HPPVSPSKTEKHSVPVPSAKTERHSPASSSSKTEKH-----SPVSPS--TKTERHSPVS 1689
QY 386 ---KGGPPAKPKGKSHDK--PLKAKGRSGILNGQKSTGNSKSPKGLKTPKTKMQMTL 440
DB 1690 TKTERHPPVSPGK--TKKPPVSPSGTE-----KHP-----1721
QY 441 LDMAGTQKMTRAPRNSGGTTPR-----TSSKPHKHLPPAALHLIAYYKENKDRKRSAL 495
DB 1722 VSPGRTEKRLPVSP--SGRTDKHPVSTAGKTEKHLVPSP-----SGKTEKQPPV 1769
QY 496 SCVISKTALLSSEDARLPEELRSVLQRYELLEHKRWASMSERQREYLLKGRLELK 555
DB 1770 S-PTSKTERI-----EETMSV-----RELN--KAFQSGQDPKSKHTGLFEHKSQK 1812
QY 556 KKLKKAERKEKE---MLERLEKQRYEDQELTKNLPAFLVDTPEGLPNTLFG---- 608
DB 1813 KQPEKGRVREKEKGPILTOREAKTENQTIKRGORLPVTGTAEKRGVRVSSIGVKE 1872
QY 609 DVAMVVEFLSCYSGLLPDAQYPIATVSLMEALSADKGGFLYLNRVLVILLQTLQDEIA 668
DB 1873 DAAGKEKVLSHK--IPE---PVQSVPEEESHRESE-----VPKEKA 1910
QY 669 EDYGEKMLKSELPLTLHVSVELVRLCLRSDD-----VOESEGSDTD-----DNKSA 717
DB 1911 DEQGDMDLQISPRKTSITDFSEVIKQLENDNDKYQFRLSEETEKALQLDQVLTSFPNT 1970
QY 718 APEDNEVQDEFLEKLETSFEFFELTSEKQ---ILTALCHRLMTYSVODHMETROOMSA 774
DB 1971 TPLDITMKDEFPLALSLQSGALDGSSESUKNEGAVGSPGSL-----MEGTPOISS 2021
QY 775 E--LWKERLAVLKEENDKKRA-EKQKRKEMAEKNGKENGVLGKTDKRGKRVKPEQV 831
DB 2022 EESYKHEGLAETPETSPLSPKSEETGTETKTE---TTTEIRSEKHP-- 2074
QY 832 DTEAEDMISAVKRRLLATQAKEREIQBERMKVLERQAEERIRKHAAKAFQEG 891
DB 2075 --TTKIDITGSEBEGATVTE-----DSETSTSEFQKATLGLSPKDTSPKQDDCT 2122
QY 892 AKAKLVM--RRTPLGTDRNHRNRYLFSDEVPLFIEKGWVHDSIDYRPNHCKDHTVSGE 950
DB 2123 GSCSVALAKETPLGLTEE-----AACDEGORTFGSS--AHKT---QTDSEVOESTATSD 2172
QY 951 DYCPRSKANLGNASMTQHGATVAVETTPKQGNLWFLCDSQKEL-----DELLNC 1006
DB 2173 -----TKALPLPEASVKTDTGTESKPOGVIRSP--QGLELALPSRDSVLSAVADSLAV 2225
QY 1007 LH-----POGIRESQLKERLEKRYQDIHSHIHLARKPNLG-----1041
DB 2226 SHKDSLEASPVLEDNDSHKTPDLSPLK---ESPGRDSLESSPPEPKMAGIFPSHP 2282
QY 1042 LKSCDGNQBLNLFRLDLVATRLQKGLGYVEETSEFEARVISLEKLKDFGECVIALQ 1101
DB 2283 LPAAVAKTELL-----TEVASVRSLRLRDPGSAEDD-----SLE-----Q 2318

QY 1102 ASVIKKFLQGFMAP-KQKRKLQSEDSAKT-----EVDEEK-----1137
DB 2319 TSLMESSGKSPSPOTPSSEVSEYVTKTDTVSTPKPAVHCEABEDDSENGINEKFRFTP 2378
QY 1138 -----KWEAEKVASALEK-----WKTAREAQTFSRMHVLLGLMLDACIKWDMASANA 1185
DB 2379 EEMFMVWTKIKWFELEQEAQKQDYKKEPKQSESS-----SSDPDA 2422
QY 1186 RKVC---PK---KGEDDK---LILCDECNKAFHLFCLRPALYVEPDG-----E 1225
DB 2423 DCSVDVDEPKHTSGSEDESGVPVLTSESRKVSSESEPELAQLKKGADSGLLPEPVIR 2482
QY 1226 WQCPACQATARNR-----GRNYTESASEDEDESDESEEEEEEEEDY 1275
DB 2483 VQPPSLPSSMDNSNPEVQFPVVVKQYTFK--MNEDTQEPGKSEEEKSLEHAEQ 2541
QY 1276 EVAGLRLRPRKTIRGKHSVIPPAARSGRRPGKPHSTRSQKAPPVDDAEVDDELVLQTK 1335
DB 2542 LAREL-----DFTBEQIHQIRIENP 2561
QY 1336 RSRRSQSLQKCEBILHKIVKRYFSWPPR-----BPVTRDEADYDVITHPM 1384
DB 2562 NSLQDS-----HALLKY---WLERDGKATDTNLVECLTKINMD-----IVHLM 2604
QY 1385 DFQT--VQKSCGYSYRSVQEFLLT--DMKQVFTNAEYVNCRGSHVLSMWKTEQCLVLLH 1441
DB 2605 ETNTEPLQBRIS--HSYAEIEQITILDHSEGFSLQBELCTAQH-----KQKEQAVSKE 2657
QY 1442 KHLPGHPYVRKR-----KKFPDLAEDEGSEPEAV--GQSRDRDRSREAEIQEWLQ 1493
DB 2658 SETCHPPIVSEBISVGYSTFDQGVPKTEGDSATLFPQTHKEQVQDDFGKMQDLPE 2717
QY 1494 DTSL 1497
DB 2718 ESSL 2721

RESULT 46

US-09-949-016-7660
; Sequence 7660 Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 7660
; LENGTH: 2753
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-7660

Query Match 3.4%; Score 272; DB 4; Length 2753;

Best Local Similarity 19.4%; Pred. No. 3.6e-10;
Matches 327; Conservative 247; Mismatches 606; Indels 504; Gaps 80;

QY 19 EEPFFTIPTQFAFRTREEYEARLRYSE--RIWTC-KSTGSSQLTHKAWBEEQVAEL 75
DB 1337 EEP-----GEPFEIVERVKEDLEKVNELIRSGICTRDESSVQSSRSRGLVEEWIV 1389
QY 76 LKEEFPWTEKVLVLEWVHNTASLEKLVDTAWLEIMTKYAVGEECDFFVGEKMKLVKIV 135
DB 1390 SDEIEEARQKAPLE-----ITEY-----PC-VEVRIDKEIKGVE 1424

Db 1684 LSSEBQELCLRKQYLSNEMQOQKUTSVTLEMEKLAABKQTEQLSELEVARLO-LQ 1742
QY 142 KVDEEA-----TEKSDGACDPSKSDKENSQIAQDHQKXETVVKEDGREESIN--- 191
Db 1743 GLDLSRSLAGIDTDAIQGRNESCDISKEHTSETTTPKHVDHICDQKDAQDNLNDI 1802
QY 192 DRARRSPKPLTSLKGER-----KWAPPKFLPHKYDVKLQNEKDISN-----VPA 238
Db 1803 EKITETGAVPTGECSEGOQSDPTNYEPPG-----EDKTQGSSECSISELSFGPNALVPM 1856
QY 239 DSLIRTERPKNKEIVRFIRHN---ALRACTGENAP---WVEDELVKYSLPKF-----S 289
Db 1857 DFL-----GNQSDI-----HNLQLRVKETSNENLRLLHVIEDRRKVESLNMKELDS 1905
QY 290 DFLDPPYKMYT-----LNPSTKRKNTGPDPRKPSKSKTNDSSL-----SPLN 333
Db 1906 KLHLQEVQMTKIEACIELEKIVGELKENSIDSEK---LEYFSCDHQELLQRTVETSEGLN 1963
QY 334 PKLWCHV---HLKXSLGSPKLVKNS-----KNSKSPPEH-----LEEMMKM 372
Db 1964 SDLEMHADKSSREDIGNAVANDSWKERFLDVENELSRIRSEKASTEHEALYLEADLEV 2023
QY 373 MSPNKLHTNPHI PKGPPAKPGKSHDKPLKAKGRSGKILNGKSTGNSKSPKGLKTPK 432
Db 2024 VOTEKL-----CLEKONENKQKIVCLBEELS VVTSESNQRLGELDTMSKKT---TALDQLS 2077
QY 433 TKMKQMTLMDMAGTKQWTRAPRNSGOTPRTSKPKHLPALHIAIYKKNKREDKR 492
Db 2078 EKKKEK-----TQEL-----ESHQ---SECHCIQV---ABAEVKET 2109
QY 493 SALSCWISKARTLLSSEDRARLPEELSLVKRYEL-----LEHKRWASMSBQRKEYLK 548
Db 2110 ELLOTLSSDVSELL---KDKTHLEKQLQSEKDSQALSILTKCELENQIAQLNKE---KELLV 2165
QY 549 KKEEELKKLKE-----K 561
Db 2166 KESELOARLESSEDEYKIANVSKALEAALVEKGBFALLRSLSTQBEVHQLRGIEKLRVRIE 2225
QY 562 AKERREKEMLERLEKOKRYEDQELTGKLPAPFLVDTPEGLPNTLFGDVAVWVEFLSCYS 621
Db 2226 ADEKKQLHTAEKUKERENDS-----LKKVENLEREL-----QMEENQ 2266
QY 622 GLLLPDAQPYITAVSLMEALSADKGGFLYLNRLVILLOTL-----LODEIAEDYGEIG 675
Db 2267 ELVILDAENSKAEVETLKTQIEEMARSL---KVFELDLVLRSEKENLTKIQIOEQGO-- 2321
QY 676 MKLSEPLTHSVSELVRLCLRRSDVQ---BESGGS-----708
Db 2322 --LSELDKLLSFKSLLEE-KEQAEIQIEESKTAVEMLQNLKELNEAVALCGDQEI 2378
QY 709 -DTDDNKDSAAFDNEVDQEFLEKLETSFEFFELTSEKLIQILTALCHRIILMTYSVODHME 767
Db 2379 KATEQSLDDPIIEEHQLRNS-IEKJRA-----LEADEKKQ-----LC--VLQJKESEH-- 2425
QY 768 TROQMSAELWKERLAVLKEENDKRAEKOKRKEAMEAKNKGKVENGLGKTRDKRIVK- 826
Db 2426 -----HADLLKGRVENLERELEART-NQEHAALEAENSK-GEVETLKAKIEGWTQSLRG 2478
QY 827 FEPQVDT---EADMTISAVKSR---LLAIOAKKEREIQERE-----MKVK-----866
Db 2479 LEDDVVTVIRSEKEDLTNELQEQERISELEIINSFENILQEQEKQVKQMEKKSSTAMEM 2538
QY 867 LERQAE--ERIR--KHKAABAFQEGIA-----KAKLVMTPTIGTDRNHNRYW 913
Db 2539 LQTLKELNERNVAALHNDQACAKQONLSSQVECLELEKAQLLQ-----GLDEAKNNYI 2593
QY 914 LFSDEVPGLFIEKGWVHDSIDYFNHHCQDHTVSGDEDCYPRSKKANLGNASMTQHTG 973
Db 2594 VLOSSVNL-----TQVEDGKQLEKDEEISRLKNQI-QDQEQIVSKLSQVEGEHQL 2646
QY 974 ATEVAVETTPKQGNLWFLCDSQKELDELNLNCLHPQGIRESQJERLEKRYQDIHTSH 1033

Db 2647 WKE-----QNL-ELRNLTVELEOKIQVLOSKNASLQDTLTVLOSSYKNLENELE 2694
QY 1034 LARKPNLGLKSCDGNQELLNFLR--SDLIEVATRLQKGGIYVEETSEFARVISLKLK 1091
Db 2695 LTK-----NDKMSFVEKVNKMTAKETELQREHMAQKTAELQEE--LSGEKNR 2741
QY 1092 DFQECVIALQASVITKFLQGFMAPKOKRRKJQSEDSAKTEEVD--BEKQWVEBAKVASAL 1149
Db 2742 LAGE-----LQLLLEIKSSKQDLKELTLENSELKXSLDCMHKQDQVEKEGKVREEI 2792
QY 1150 EKWKTAIRAQTFSRMHVLLGLDACIKWDMSAENARCKVCPPKGGEDDKLILCDEC--NK 1207
Db 2793 AAYQLLHEAE---KKHQAL--LLDTNKQYVEVEIQTYR-----EKUTSKEECLUSSQ 2838
QY 1208 AFHLFCRLPALYEVDPGEWQCPACQAPATARRNSGRGNYTEESASESEDD-----ESD 1260
Db 2839 KLEIDLKSSKEELNLSKATTOILEELKKTMDNLKYVNLKKNENAKGKMLLKSC 2898
QY 1261 EEEEEEEEEEDYEVAGLRURPRKTIRGKHGSHVIPPAAARSGRRPGKKPHSTRRSQPKAP 1320
Db 2899 KQLEEEKEILQKELSOL-----QAAQEKQKQTG-----2925
QY 1321 PVDAAVDELAVLTQKSSRRQSLQKCEILHKIVKYRFSWFPFVTRDEADYD-- 1378
Db 2926 TVMDTKVDEL-----TTEIKELKETEELTK-----EAEYLDKY 2960
QY 1379 ---VITHPMDFTQVQNKSCGYSRVSQEF--TDMKOVFTNAEYVNCRGSHVLSVMKTEQ 1434
Db 2961 CSLLSISHE-----KLEKAKEMLETOVAHLCSQSQKQDSRGSGPLGPPV--- 3003
QY 1435 CLVVLHLHGLPHG---PYVRRKRKKFPDRLAEDGDESPQVQSGDREDRRSREARIQE 1490
Db 3004 -----PGSPPIPSVTEKRL-----SSQNKASGKRQSSSGIWE 3036
RESULT 48
US-09-538-092-1078
; Sequence 1078, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
; APPLICANT: Glot, Loic
; APPLICANT: Mansfield, Traci A.
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538,092
; CURRENT FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CuraPatSeqFormatter Version 0.9
; SEQ ID NO 1078
; LENGTH: 1976
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: Polypeptide Accession Number P35580
US-09-538-092-1078

Query Match 3.3%; Score 265; DB 4; Length 1976;
Best Local Similarity 18.9%; Pred. No. 7.1e-10;
Matches 335; Conservative 263; Mismatches 566; Indels 606; Gaps 81;
QY 15 PLFGSEPFPTIHTQEAER---TREEYEALR-----BRYSERIWTCKS 54
Db 316 PIPGQDKDNFOETMEAHMFGSHIEILSMKLVSSVLQFGNISPKKERNTDQASMPEN 375
QY 55 TGSQSLTH-----KMAWEEQ---EVAELLKKEFFAWYK 86
Db 376 TVAQKLCHLLGMVMEFTRAILTPRIKVGDRVYVQAQTKQADFAVEALAK----ATYER 431

```

QY 87 LVLEWVHNTASLEK-----LVDTAWLEIMTKYAGBECDFEVCGERKMLKVIK 136
Db 432 LFRWLVRINKALDRTKROGASFIGLDIAGFEI-----FELNSFEOL----- 474
QY 137 IHPLEKVDEATEKKSADGACDPSDDKENSIOAQDHQKKTETVWKEDDEGRRESINDRARR 196
Db 475 -----CINYNEK-----LOQLPNHMTFELQEEVQREGI----- 504
QY 197 SPRKPTSLKGERKWAPKPFHPKYDVVKVQLONEDKIIISNVPADSLIRTERPNKBEIVRYF 256
Db 505 -----EW-----NFIDFGLDLQ-----PCIDLI-----ERPANPCGLAL 534
QY 257 IRHNALRAGTGENAPWVZE-----DELAVKYSPLSKP-----SDFLLDYPKYMT 300
Db 535 L-----DEECFFPKATDKTFVEKLVQBGSHSKFQKPRQLKADFCIIHY----- 580
QY 301 LNPSTKRKNTGSPDRKPS-----KSKTDNSSLSSPLN-----PKLWCHVH----- 341
Db 581 -----AGVDYKADRWLMKNWDPNDNVNATLLHQSSDRFVAELWQDVRIIVGLDV 631
QY 342 --LKXSLGSPKLVKNKSNKSPPEEHLEEMKQMS-P-NKLHTNF-----HIPKGPAPKPG 395
Db 632 TGMTETAFGSAYKTKKGMPTVGQLYKESLTKLMATLRNTNPNFVRCIIPNH-----EKRG 688
QY 396 KXSDPLKAKGRSGILNG-----OKSTGNSKSPKPKGLTKPTKMKQMTLLDMAG 446
Db 689 KLDPLHLVDOLRCNGVLEIRICROGFPNRIVFQEFORYEILTNAIPK--GFMDGKQA 746
QY 447 TOKMTRAPNSGOTPTSSKPHKPLPPAALHIIAYVKNKREDKRSALSCVISKTAR-L 505
Db 747 CERMIPALBDNLNIRIGOSKIFFRAGVLAHL-----EEEDLKITDIIIFQAVCRC 800
QY 506 LSSEDRARLPEELSL--VQKR---YELLEHKRWASMESEQRKEYLKKRBEELKKLKE 560
Db 801 LARKAFKKQOQSALKVQLQRNCAAYLKLHQQWVRFTKVPKLQVTRQEEELQAKDBE 860
QY 561 KAKERREKEMLE-RLEKQRYEDQELTGKVLPAFLVDTPEGLPNTLFGDVANVVEFLSC 619
Db 861 LLKVKREKOTKVEGELEEMERKHQQLLEEKNI----- 891
QY 620 YSGLLLPDAQYPITAVSLMEALSADKGGFLYLNRLVLILOTLLODEIADYEGELGKMLS 679
Db 892 -----LAQQLA-----ETELFAAEEMEARLAARKQ 918
QY 680 EIPLTILHSVELVRLCLRSDVOESEGS---DTDNNDKSAAPFENEVO-DE----- 727
Db 919 ELEELIHLDL-----ESRVEEERQILQNEKKQMAHIQDLEEQDDEEGARQKL 969
QY 728 FLEKLETSFFELTSEKLOIL-----TALCHRIILMTYSVDHMETRQMSAEL 776
Db 970 QLEKV-TABAKTKMVEEILLLEDQNSKFIKEKKLMEDRIACSSQLAESEKAKNLAKI 1028
QY 777 WKERLAVLKEENDKKAERKQKEME-AKNKENGKVENGGLTKDRKRVKPEPOVDTEA 835
Db 1029 RNKQEWISDLERLKEBKTQLEKAKRLDGET-----TDLODQIAELOAQIDE-- 1080
QY 836 EDMISAVKSRRLIAIQ-AKKEREIQ-----EREMKVKLERQAE-----BERIR 877
Db 1081 -----LKLQAKKEEELQALARGDDETLHKNNALKVVRVRELQAIQAEQLQEDFE 1128
QY 878 KHKAAAEKAFQEGIAKAKLWMRTPIGTDNRNHNRYLFSDEVPGLPFIEKGWHDSDIDYRF 937
Db 1129 SEKASRNKAEKQ---KRDLE-----SELEALKTE---LEDTLD--- 1160
QY 938 NHHCKDHTVSGDEDYCPRSKK-----ANLGN-----ASMNTQHTATEVAVET 981
Db 1161 -----TTAAQQL--RTKREGEVLELKALEETKNHEAQIQDMRQRHATALELSQ 1211
QY 982 TTP-----KQGNLWFLCDSQKDELINCLHPQGIRESQKLERKRYODIITH 1030
Db 1212 LEQAKRFKANLEKNKQ-----LETDNKELACEVKVLQVQKAESEHKRKLDAQVE--- 1263

```

```

QY 1031 SIHLARKPNLGLKCDGNOELLNPLRSDLIEVATRLQ-----KGGLYGVEE 1076
Db 1264 -----LHAKVSEGR-----LRVELAKASKLQNELDNVSTLLBEAEKGIKFAKD 1309
QY 1077 TSEFARVISLEKLKDFGECVIALQASVIFKLOQFMAPKQKRRKLQSEDSAKTEVEBDE 1136
Db 1310 AASLESQLODQTEL-----LQETRQK-----LNLSSRIQLBEEKNSLQEQESEE 1355
QY 1137 KMWTEAKVASALEKWKTAIRB--NOTFSRMHVLLGMLDACIKWMDSAENARCKVCPKKG 1194
Db 1356 -----BEAR--KNLEKQVLAQLADTKKQVDDDLGTIE-----SLEAAKXGCLL-KDA 1401
QY 1195 E-----DDKILDECNKAFHLFCLRPALYEVF--DGEWQCPA-----CQPATAR 1237
Db 1402 EALSORLEEKALAYDKLEKTNR--LOQELDDLTVDLHQROVASNLEKKQKFPOLLAE 1459
QY 1238 RNSGRNRYTESASDESDDESEDEEE-----EEEEEEEDYEVAGLRLP--RKT 1287
Db 1460 EKSISARYAEE--RDRAEAERAREKETKALSARALEEALEAKEFERQNKQLRADMEDL 1516
QY 1288 IRGKHSVTPPAARSORRPGKPHSTRSOPKAPPVDDAEVDELVLQTKRSSRRQSLELOK 1347
Db 1517 MSSKDDV-----GKNVHELEKSKRAL-----EQVEEMRTQLEELED-----ELQA 1557
QY 1348 CE-----EILHKIVKRYFSPFPREPVTDRDAED-----YYDVITHPMDFOQVQNK 1393
Db 1558 TEDAKLRLEVNMQAQAEF--RDLQTRDEQNEEKKLLIKQVRELAELEDERKQRAL 1614
QY 1394 SCGSYRSVQEFITDMKQVFTNAEVYVNCRGSHVLSM----- 1429
Db 1615 AVASKKME--IDLKLEAQIEAANKARDEVIKQLRKLQAKMDQVQRELEEARASRDEI 1671
QY 1430 -----VKTEOCLVLLHKLPHGPHYVRRKPKFPDRLAEDEGSEPEAVQGSR-- 1477
Db 1672 FAQSKESEKKLSELAELQLQEEELASSERARRHAEQERDELADEITNS---ASGKSALL 1728
QY 1478 DEDRE--SREAEIQEWLODTSLYSAKINSK 1505
Db 1729 DEKRREARIAQLEEELEEBQSNMELNDR 1758

```

RESULT 49

```

US-08-353-700-1
; Sequence 1, Application US/08353700
; Patent No. 559919
; GENERAL INFORMATION:
; APPLICANT: YEN, TIMOTHY J.
; APPLICANT: RATTNER, JEROME B.
; TITLE OF INVENTION: NUCLEIC ACID ENCODING A
; TITLE OF INVENTION: TRANSIENTLY-EXPRESSED KINETOCHORE PROTEIN,
; TITLE OF INVENTION: AND METHODS OF USE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSER: DANN, DORFMAN, HERRELL AND SKILLMAN
; STREET: 1601 MARKET STREET, SUITE 720
; CITY: PHILADELPHIA
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/353,700
; FILING DATE: 09-DEC-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: REED, JANET E.
; REGISTRATION NUMBER: 36,252
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 563-4100

```



```

; TELEFAX: (215) 563-4044
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3248 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: HUMAN
; US-08-353-700-1

Query Match 3.3%; Score 264; DB 1; Length 3248;
Best Local Similarity 18.1%; Pred. No. 1.7e-09;
Matches 305; Conservative 289; Mismatches 570; Indels 524; Gaps 76;

QY 52 CKSTGSSQTHKE-----AWEEQEVALLKEEPPAWYKLVLR--MVHNTASLEKL 102
DB 1624 CSSLQENLTRKETPSAPAKGVBEESLCEVYRQSLKLEEKMESOGIMKNKEIQLEQL 1683
QY 103 VDT-----AWLEIMTKYAVGEECDPEVGKEK-----MLKVKIVKIHPL 141
DB 1684 LSSEQLDCLRKQYLSSENEQWQOKLTVTLEMESKLAEEKQTEQLSLEVARLQ-LQ 1742
QY 142 KVDEEA-----TEKSDGACDPSKSDKENSQIAQDHQKKEVTVKDEGRRRESIN--- 191
DB 1743 GLDLSRSLGIDTEDAIQRNESCDISKEHTSETTERTPKHVDVHOICDKDAQDNLNDI 1802
QY 192 DRARRSPRLPTSLKGER-----KWAPPKFLPHKYDVKLQNEDKIISN-----VPA 238
DB 1803 EKITETGAVKPTCEGSGEQSPDNTYEPGP-----EDKTQSGSECISELSFSGPNALVPM 1856
QY 239 DSLIRTERPNKEIVYFIRHN---ALRAGTGENAP--WVVEDELVKYSLSKPF-----S 289
DB 1857 DFL-----GNQBDI-----HNLQVRKETSNTSNTLHVLVIEDRDKVESLNNMKELDS 1905
QY 290 DFLDLPKYMT-----LNPSTKRKNTGSPDRPKSKSKTDNSSL-----SSPLN 333
DB 1906 KLHLQEVQLMTKIEACIELEKIVGELKENSIDLSEK--LEYFSCDQHELLQRVETSEGLN 1963
QY 334 PKLWCHV--HLKSLGSGPLKVNKS-----KNSKSPPEH-----LEEMKXN 372
DB 1964 SLEMHADKSSREDIGDNVAKVNDKWKERFLDVENELSRSEKASIEHAELYLEADLEV 2023
QY 373 MSPNKLHTFHPKPKPPAKPKGKSDKPKAKGRSGKILNGOKSTGNSKSPKGLKTPK 432
DB 2024 VQTEKL-----CLEKDNENKQKIVIVCLEBELSVVTSERNQLRGELDTMSKKT--TALDQLS 2077
QY 433 TKMKQMTLLDMAKGTOKMTRAPRNSGGTPTRTSSKPHKHLPPAALHLIAVYKKNKDRDXR 492
DB 2078 ERMKEK-----TQEL-----ESHQ-----SECHCIQV--AAEAIVEKKT 2109
QY 493 SALSVCVISTARLLSSEDRARLPEELRSIVQRYEL-----LHKKKWASMBEQRKEYLK 548
DB 2110 ELLQTLSSDVSELL--DKNTHLQKLEQLSKLEKQALSLSLTKCELENQIAQLNKE--KELIV 2165
QY 549 KKEELKKLKE-----K 561
DB 2166 KESESLQARLSSESYDEKLVNASKALEAALVEKGFALRLSSTOBEVHQLRRGIEKLRVRIE 2225
QY 562 AKERREKEMLERLEKOKRYEDQELTKNLPAPFLVDTPGLPNTLPFGDVAMVVEPLSCYS 621
DB 2226 ADEKQLHHAETKUKERENDS-----LKDQVENLREL-----QMSERNQ 2266
QY 622 GLLLPDAQPIPTAVSLMEALSADKGGFLYNRLVILQV-----LLQDEIAEDYGEIGMKL 678
DB 2267 ELVILDAENSKAEVETLTKQIEBARSLSKIFELDLVTLRSEKENTLQIQEKGQO-----L 2322
QY 679 SEIPLTLHVSVELVRLCLRRSDVQ--EESSEGS-----DT 710
DB 2323 SELDKLLSLFKSLLEK--KEQABTIQKEESKTAVEMLQNLKELNEAVALCGDQETMKAT 2381

```

```

QY 711 DDNKDSAAFEDNEVQDEFLEKLETSSEFFELTSEKLIQILTALCHRLIMTYSVODHMETRQ 770
DB 2382 EQSLDPPIEEHQLRNS-IEKLRA-----RLEADEKKQ-----LC--VLQKLESEH----- 2425
QY 771 QMSAELKWERLAVLKEENDKKABKQKREMAKANKENGKVENGLKTDKRGKRVKPEPQ 830
DB 2426 --HADLLKGRVENLRELEIART-NQEHAALEAENS-K-GEVET----- 2464
QY 831 VQTEAEDMISAVKSRRLLAIAQAKEREIQEREMKVKLER-----QAEEE 874
DB 2465 LKAKIEGWTQSLRGLELDVVTIRSEKENUTNELQKEQERISELEIINSSFENILQKEQE 2524
QY 875 RIR---KHKAAAE-----KAFQEGIA-----KAK-----LVMRRTPI--G 904
DB 2525 KVMQKEKSTAMEMLTQQLKELNERVAALHNQDEACKAEQNLSSQVECELEKELKALQQLQ 2584
QY 905 TDRNHNRYWLFDEVPGLFIEKGWVHDSIDYRPNHCKHKOHTVSGDDSDYCPRSKANKLGN 964
DB 2585 LDEAKNNYIVLQSSV-----KGLIQEVEDGKQLEKDEEISRLKNQI-QDOEQLVSKL 2637
QY 965 ASMNTQHGATATVAVETTTTPKQGNLWFLCDQKELDELLNCLHPQGIRESQLEKRLER 1024
DB 2638 SQVEGEHQLWKE-----QNL-EURNLTVELEQKIQVLQSKNASLQDTLTVLQSS 2685
QY 1025 YQDIISHLARKPNLGLKSCDGNQOELLNFLR--SDLIEVATRLQKGGLYVEETSEFEA 1082
DB 2686 YKNLENELETK-----MDKMSFVEKVNKWTAKETELQREHMAQKTAELQ 2733
QY 1083 RVISLEKLDKFGECVIALQASVVKPLQGFMAPKQKRLKQSDSDAKTEVD--EBKMW 1140
DB 2734 E-LSGKKNRNLAGE-----LQILLLEIKSKQQLKELTLENSLKKSLDCMHKDOVE 2783
QY 1141 BEAKVASALEKWKTAIREAQTFSRMHLVGLMLDACIKWMSAENARCKVCPKKGEDDKLI 1200
DB 2784 KEGKVRBEETAYQLRLHEAB---KKHQA-LLDTNKQYEVETQTYR-----EKL 2829
QY 1201 LCDEC--NKAFLFLCLRPALYVPPDGEWQCPACQATARNRSGRNYTBESASEDSED 1257
DB 2830 SKRECLSSQKLEIDLLKSKSEELNLSLKATTOILEELKTKMDNLYVNLKKNERAQG 2889
QY 1258 -----ESDEEBEEREBEEDYEVAGURLRPRKTIRGKHSVIPPAASGRRRPKKPHS 1311
DB 2890 KMKLLIKSKQLEEBEKEILQKELSQ-----QAQEKQKTG----- 2925
QY 1312 TRSQPKAPPVDDAEVDELVLQTKSSRRRQSLQKCEILHKIVKYRFSWPPEPVTRD 1371
DB 2926 -----TVMDTKVEL-----TTEIKELKETLBEKTK----- 2951
QY 1372 EAEDYD-----VITHPMDFTQVQNKSCGYSYRSVQEF--TDMKQVFTNAEVYVNCRGSHV 1425
DB 2952 EADEYLDKYCSLLISHE-----KLEKAKEMLETOVAHLCSQSKQSDSRGSP 2998
QY 1426 LSCMYKTEQCLVLLHKLHPGH---PYVRKPKKFPDLRAEDGSEPEAVGQSDRDR 1482
DB 2999 LGPVV-----PGSPSPSVTEKEL-----SSGQNKASGR 3028
QY 1483 SREASIQE 1490
DB 3029 QRSSGIWE 3036

```

```

RESULT 50
PCT-US95-16216-1
; Sequence 1, Application PC/TUS9516216
; GENERAL INFORMATION:
; APPLICANT: Yen, Timothy J.
; APPLICANT: Rattner, Jerome B.
; TITLE OF INVENTION: Nucleic Acid Encoding a Transiently
; TITLE OF INVENTION: Expressed Kinetochores Protein, and Methods of Use
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dann, Dorfman, Herrell and Skillman

```

```

; STREET: 1601 Market Street Suite 720
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/16216
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/353,700
; FILING DATE: 09-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Reed, Janet E.
; REGISTRATION NUMBER: 36,252
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 563-4100
; TELEFAX: (215) 563-4044
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3248 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; PCT-US95-16216-1

Query Match 3.3%; Score 264; DB 5; Length 3248;
Best Local Similarity 18.1%; Pred. No. 1.7e-09;
Matches 305; Conservative 289; Mismatches 570; Indels 524; Gaps 76;

QY 52 CKSTGSSQLTHKE-----AWEEOEVAELKEEFPWYKLVLE--MVHNTASLEKL 102
DB 1624 CSSLQENLTKETPSAPAKGVVEELCEVYRQSLKLEERWESQGIMKNKEIOLEQL 1683
QY 103 VDT-----AWLEMTKYAVEECDFEVGKEK-----MLKVKIVKIHPLE 141
DB 1684 LSSEQELDLCKRYLSENEQOQKLTSTLEWESKLAEEKOTQOLSLELEVAKLQ-LQ 1742
QY 142 KYDEBA-----TEKSDGACDSPSDKENSQIAQDHQKKEITVVKEDGRRRESIN--- 191
DB 1743 GLDLSRSILGIDTDAIQGRNESCDSKEHTSETTPKHVDVHQICDXDAQDNLNDI 1802
QY 192 DRARSPKLPSTLKKGER-----KWAPPKFLPHKYDVVKLQNEKDITISN-----VPA 238
DB 1803 EXITETGAVKPTGECGSEQSPDTNVEPPG-----EDKTQGSSECTISELSFSGPNALVPM 1856
QY 239 DSLIRTERPKNKEIVYFRHN-----ALRAGTGENAP--WVVEDELVKYKSLSPKF-----S 289
DB 1857 DFL-----GNOEDI-----HNLQVRKETSINLRLHLVIEDRDRKVESLLNEMKELDS 1905
QY 290 DFLDLPYKYMT-----LNPSTKRKNKTSQPKPSKSKTQNSL-----SSPLN 333
DB 1906 KLHLQEVQLMTKIEACIELEKIVGELKKNESDLSEK--LEYFSCDHQELLQRVETSEGLN 1963
QY 334 PKLMCHV--HLKKSLSGSPKVKNS-----KNSKSPBEH-----LBEEMKQ 372
DB 1964 SDLEMHADKSSREDIGDNVAKVNDWSKWERFLDVENELSRIRSEKASIEHEALYLEADLEV 2023
QY 373 MSPNKLHTNFHPKKGPPAKPKGSKDKPLKAKGRSKGLNGKSTGNSKSPKKGLKTPK 432
DB 2024 VQTEKL-----CLEKONENKQKIVIVCLEEELSUVVTSEBRLGELDTMSKKT--TALDQLS 2077
QY 433 TKMKQWTLDMAGTKQKMTAPRNSGGTPTRTSKPKHKLPPAALHLIAYIKENKQREDKR 492
DB 2078 ERMKEK-----TQEL-----ESHQ-----SECLHCIVQ--AAEAVKEKT 2109
;
493 SALSVCVISTARTALLSSEDRARLPBELRSIVOKRYEL-----LEHKRWASMSSEORKEYLK 548
2110 ELLQTLSSDVSELL--KDKTHLQEKIQLSEKDSQALSITKCELENQIAQLNKE--KELLV 2165
QY 549 KKREELKKKLE-----K 561
DB 2166 KESESLOARLSSEDYKELNVSKALEAALVEKGEFALRLSSTOEEVHQLRGIEKLVRVE 2225
QY 562 AKERREKEMLEKLEKQKRYEQDELTKNLPAPRLVDTPEGLPNTLFGDVAMVVEFLSCVS 621
2226 ADEKKQLHIAEKLERERENDS-----LKDQVENLEREL-----QMSEENQ 2266
QY 622 GLLLPDAQYPITAVSLMEALSADKGGFLYLNRLVILLQT---LLQDEIAEDYGELGMKL 678
2267 ELVILDAENSKAEVETLTKQIEEMARSLKIFELDLVTLRSEKENITKQIOEKQGG---L 2322
QY 679 SEIPLTLHSVSELVRLCLRRSDVQ--EESSEGS-----DT 710
2323 SELDKLLSFPKSLLEE--KEQAEIQIKEESKTAVEMLQNLKELNEAVALCGDQETMKAT 2381
QY 711 DNKDSAAFPENEVODEFLEKLETSFEFFELTSEKQLQILTALCHRLIMTYSVODHMETRQ 770
2382 EOSLDPPIETEEHQLRNS-IEKLRA---RLEADEKKQ---LC--VLQOLKESEH----- 2425
QY 771 QMSAELWKEKRLAVLKEENDKKRAEKQKREKEMKENGKENGKGLTKDRKKRIVKFPFEP 830
2426 --HADLLKGRVENLERELEIART--NOEHAALAEANSK-GEVET----- 2464
QY 831 VDTEAEDMISAVKSRRLAIQAKKERETQEREMKVKLER-----KAK-----LVMRRTPI--G 904
2525 KVQMEKESSTAMEMLOTQKELNERVAALHNDQEAQKAKQNLSSQVECLEKEKALQLOG 2584
QY 905 TDRHNRYWLPSPDEPGFLIEKGWVHDSIDYRFNHHCKDHTVSGDEDYCPRSKANLGN 964
2585 LDEAKNNYIVLGSSV-----KGLIQEVEDGQKLEKDEEISRLKNOI-QDOQLVSKL 2637
QY 965 ASMTQHGTADEVAVETTTTPKQGNLWFLCDSQKELDELLNCLHPQIGRESQKLERLEKR 1024
2638 SOVEGEHQLWKE-----QNL-ELRNLTVELEQKIQVLQSKNASLQDTLTVLQSS 2685
1025 YODIITHS IHLARKPNLKGSCDGOELLNFR--SDLIEVATRLQGGKGYVEETSEFPA 1082
2686 YKNLENELELTK-----MDKMSFVEKVNKMTAKETELQREHMAQKTAELQOE 2733
QY 1083 RVISLEKLDKFGECVIALQASVIKKFLOGFPWAPKQKRRKLQSEDSAKTEEVD--EKKNV 1140
2734 E-LSGEEKRLAGE-----LQLLLEIKSSKDQLBELTLENSELKSLDCMHKQOQE 2783
QY 1141 BEAKVASALEKWKTAIREAQTFSRMHVLLGLMDACIKWDMSAENARCKVCPKKGEDDKLI 1200
2784 KEGKYREBIAEYQLRLHEAE---KXQAL-LLDTNKQYEVEIQTYR-----EKL 2829
QY 1201 LCDEC--NKAFHFLCLRPALEVPDGEWQCPACQATARNRSGRNYTEESASESDED- 1257
2830 SKEECLSSQKLEIDLLKSKSEELNLSKATQILBELKKTMDNLKYVNLKKNENARAQ 2889
QY 1258 -----ESDEEEEEEEEEEDYEVAGLRLRPRTIRGKHSVIPPAARSGRRPGKPKHS 1311
2890 KMKLLIKCKQLQEEKEILOKELSOL-----OAOEKKQTG----- 2925
QY 1312 TRRSQKAPPVDDAEVBLVLTQTKRSSRRQSLEOKCEBILHKIVKIRPSWFPFVTRD 1371
2926 -----TVMDTKVDEL-----TTEIKELKELTEKTK----- 2951
QY 1372 EABDYD-----VITHPMDFQTVQNKSCGYSYRVOEFL-TDMKQVFTNAEYVNCRGSHV 1425
2952 EADEYLDKYCSILLISHE-----KLEKAKEMLETQVAHLCSQSQSDSRGSP 2998

```


THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 9, 2005, 13:18:56 ; Search time 177.909 Seconds
(without alignments)
5405.490 Million cell updates/sec

Title: US-10-702-148-13
Perfect score: 9826
Sequence: 1 MEMEANDHFNFTGLPPA.....MRRPFESRWSEFYQKQANL 1878

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 65 summaries

Database : UniProt_03:
1: uniprot_sprot:
2: uniprot_trembl:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	9716	98.9	1878	1 BA2A_HUMAN	Q9uif9 homo sapien
2	9571.5	97.4	1905	2 Q68DI8	Q68di8 homo sapien
3	7795	79.3	1850	1 BA2A_MOUSE	Q9vi95 mus musculus
4	3777	38.4	1698	2 Q6YI94	Q6Yi94 xenopus lae
5	3762	38.3	886	2 Q8QU42	Q8Qu42 mus musculus
6	3198.5	32.6	769	2 Q8CGH2	Q8Cgh2 mus musculus
7	2960.5	30.1	709	2 Q80VL8	Q80Vl8 mus musculus
8	2496.5	25.4	1972	1 BA2B_HUMAN	Q9uif8 homo sapien
9	2387.5	24.3	2130	1 BA2B_CHICK	Q9del3 gallus gall
10	1245	12.7	1369	2 Q69Zi8	Q69zi8 mus musculus
11	1144	11.6	2060	2 Q7Q3S9	Q7q3s9 anopheles g
12	1108.5	11.3	796	2 Q8NC87	Q8nc87 homo sapien
13	1095.5	11.1	3080	2 Q9V602	Q9v602 drosophila
14	1084	11.0	3109	2 Q9BMOQ	Q9bmq0 drosophila
15	863.5	8.8	1376	2 Q23590	Q23590 caenorhabdi
16	836	8.5	181	2 Q8BRP6	Q8brp6 mus musculus
17	736	7.5	162	2 Q8FE75	Q8fe75 mus musculus
18	660.5	6.7	1556	1 BA1A_HUMAN	Q9nr12 h bromodoma
19	538	5.5	1479	1 BA1B_MOUSE	Q9z277 mus musculus
20	529	5.4	1357	2 Q7PRP9	Q7prp9 anopheles g
21	528.5	5.4	1483	1 BA1B_HUMAN	Q9uig0 homo sapien
22	527.5	5.4	1476	2 Q9V9T4	Q9v9t4 drosophila
23	526.5	5.4	1476	2 Q9Y0W1	Q9y0w1 drosophila
24	523.5	5.3	1476	2 Q9V624	Q9v624 drosophila
25	519	5.3	1079	2 Q8UVR4	Q8uvr4 xenopus lae
26	491	5.0	1427	2 Q9N5L9	Q9n5l9 caenorhabdi
27	471	4.8	1202	2 Q7QBK2	Q7qb2 anopheles g
28	469.5	4.8	892	2 Q6P1D9	Q6p1d9 mus musculus
29	430.5	4.4	683	2 Q960I3	Q960i3 drosophila
30	425.5	4.3	718	2 Q8C8D1	Q8c8d1 mus musculus
31	407.5	4.1	861	2 Q8C0K4	Q8c0k4 mus musculus

32	406.5	4.1	2781	2 Q9UIG2	Q9uig2 homo sapien
33	396.5	4.0	627	1 BA1A_XENLA	Q8uvr5 xenopus lae
34	383	3.9	2764	2 Q7Z7D6	Q7z7d6 homo sapien
35	372.5	3.8	657	2 Q8CAU9	Q8cau9 mus musculus
36	362.5	3.7	1163	2 Q6E2N3	Q6e2n3 brachydanio
37	353	3.6	1071	2 Q8ZPX5	Q8zpx5 mus musculus
38	350.5	3.6	1209	2 Q9Z4W6	Q9z4w6 mus musculus
39	348	3.5	1214	2 Q9BQ04	Q9bq04 homo sapien
40	344.5	3.5	1051	1 TF1A_MOUSE	Q64127 mus musculus
41	344	3.5	131	2 Q8CFP4	Q8cfp4 mus musculus
42	343	3.5	1142	2 Q6SI71	Q6si71 mus musculus
43	338	3.4	1050	1 TF1A_HUMAN	O15164 homo sapien
44	335	3.4	1127	1 TF1G_HUMAN	Q9upn9 homo sapien
45	320	3.3	10820	2 Q7SDK2	Q7sdk2 neurospora
46	312.5	3.2	961	2 Q6E2N2	Q6e2n2 brachydanio
47	307	3.1	249	2 Q9ST46	Q9st46 drosophila
48	307	3.1	460	2 Q8MK29	Q8mk29 drosophila
49	307	3.1	516	2 Q8R154	Q8r154 mus musculus
50	301.5	3.1	1114	2 Q6P9L3	Q6p9l3 mus musculus
51	301	3.1	3277	2 Q7NNI4	Q7nni4 gloeobacter
52	300	3.1	2759	2 Q9VID9	Q9vid9 drosophila
53	292.5	3.0	2187	2 P70670	P70670 mus musculus
54	290.5	3.0	2768	2 Q9VC00	P70670 mus musculus
55	288	2.9	2468	1 MAPB_HUMAN	Q9vc00 drosophila
56	286.5	2.9	5412	1 FUTS_DROME	P46821 homo sapien
57	286	2.9	2464	1 MAPB_MOUSE	Q9w596 drosophila
58	286	2.9	2592	2 Q19135	P14873 mus musculus
59	285	2.9	2392	2 Q810K1	Q19135 caenorhabdi
60	285	2.9	2407	2 Q7YSH5	Q810k1 caenorhabdi
61	284.5	2.9	2805	1 MAPA_HUMAN	Q7ysh5 caenorhabdi
62	283.5	2.9	4498	2 Q93291	P78559 homo sapien
63	280.5	2.9	645	2 Q8VDN7	Q93291 figu rubrip
64	280.5	2.9	2459	1 MAPB_RAT	Q8vdm7 mus musculus
65	280.5	2.9	2774	1 MAPA_RAT	P15205 rattus norv
					P34926 rattus norv

ALIGNMENTS

RESULT 1		BA2A_HUMAN		STANDARD; PRT; 1878 AA.	
ID	Q9UIF9; O00536; O15030; Q96H26;				
AC	28-FEB-2003 (Rel. 41, Created)				
DT	28-FEB-2003 (Rel. 41, Last sequence update)				
DT	05-JUL-2004 (Rel. 44, Last annotation update)				
DE	Bromodomain adjacent to zinc finger domain 2A (Transcription				
DE	termination factor-1 interacting protein 5) (TTF-I interacting protein				
DE	5) (Tip5) (hWALP3).				
GN	Name=BAZ2A; Synonyms=KIAA0314, TIP5;				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
OX	NCBI_TaxID=9606;				
RP	(1)				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Testis;				
RA	MEDLINE=20130112; PubMed=10662543; DOI=10.1006/geno.1999.6071;				
RT	Jones M.H., Hamana N., Nezu J.-I., Shimane M.;				
RL	"A novel family of bromodomain genes.";				
RL	Genomics 63:40-45(2000).				
RP	(2)				
RP	SEQUENCE OF 332-738 FROM N.A.				
RC	TISSUE=Lung;				
RA	Jansa P., Grummt I.;				
RL	Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.				
RP	(3)				
RP	SEQUENCE OF 639-1878 FROM N.A.				
RC	TISSUE=Brain;				
RA	MEDLINE=97349984; PubMed=9205841;				
RA	Nagase T., Ishikawa K.-I., Nakajima D., Ohira M., Seki N.,				
RA	Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;				
RT	"Prediction of the coding sequences of unidentified human genes. VII.				

RT The complete sequences of 100 new cDNA clones from brain which can
 RT code for large proteins in vitro.";
 RL DNA Res. 4:141-150(1997).
 RN [4]
 RP SEQUENCE OF 1038-1878 FROM N.A.
 RC TISSUE=Lymph;
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldi M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Ustin T.B., Toehiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
 RA Bobak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S.E., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grinwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 CC -!- FUNCTION: May play a role in transcriptional regulation
 CC interacting with ISWI. May serve a specific role in maintaining or
 CC altering the chromatin structure of the rDNA locus (By
 CC similarity).
 CC -!- SUBUNIT: Together with ISWI/SNF2h, it forms a complex termed NoRC
 CC (nucleolar remodeling complex).
 CC -!- SUBCELLULAR LOCATION: Nuclear. Colocalizes with the basal RNA
 CC polymerase I transcription factor UBF in the nucleolus.
 CC -!- TISSUE SPECIFICITY: Expressed at moderate levels in most tissues
 CC analyzed, including heart, brain, placenta, lung, skeletal muscle,
 CC kidney and pancreas.
 CC -!- SIMILARITY: Belongs to the WAL family.
 CC -!- SIMILARITY: Contains 4 A.T hook DNA-binding repeats.
 CC -!- SIMILARITY: Contains 1 bromodomain.
 CC -!- SIMILARITY: Contains 1 DDT domain.
 CC -!- SIMILARITY: Contains 1 methyl-CpG-binding (MBD) domain.
 CC -!- SIMILARITY: Contains 1 PHD-type zinc finger.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AB032254; BAA9211.1; -;
 DR EMBL; AF000422; AAB50864.1; -;
 DR EMBL; AB002312; BAA20773.1; -;
 DR EMBL; BC008965; AAB08965.2; -;
 DR HSPG; Q9UG0; 1F62.
 DR Genew; HGNC:962; BA22A.
 DR MIM; 605682; -;
 DR GO; GO:0005731; C:nucleolus organizer complex; NAS.
 DR GO; GO:0003677; F:DNA binding; NAS.
 DR GO; GO:0003028; F:transcription regulator activity; NAS.
 DR GO; GO:0006338; P:chromatin remodeling; NAS.
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; NAS.
 DR InterPro; IPR000637; A.T hook.
 DR InterPro; IPR001487; Bromodomain.
 DR InterPro; IPR004022; DDT.
 DR InterPro; IPR001739; Methyl-CpG_bind.
 DR InterPro; IPR001965; Znf_PHD.
 DR Pfam; PF02178; AT hook; 4.
 DR Pfam; PF00439; Bromodomain; 1.
 DR Pfam; PF02791; DDT; 1.

DR Pfam; PF01429; MBD; 1.
 DR Pfam; PF00628; PHD; 1.
 DR PRINTS; PR00503; BROMODOMAIN.
 DR SMART; SM00384; AT_hook; 4.
 DR SMART; SM00297; BROMO; 1.
 DR SMART; SM00571; DDT; 1.
 DR SMART; SM00391; MBD; 1.
 DR SMART; SM00249; PHD; 1.
 DR PROSITE; PS00633; BROMODOMAIN_1; FALSE_NEG.
 DR PROSITE; PS50014; BROMODOMAIN_2; 1.
 DR PROSITE; PS50827; DDT; 1.
 DR PROSITE; PS50982; MBD; 1.
 DR PROSITE; PS01359; 2F_PHD_1; FALSE_NEG.
 DR PROSITE; PS50016; 2F_PHD_2; 1.
 KW Bromodomain; Coiled coil; DNA-binding; Nuclear protein; Repeat;
 KW Transcription regulation; Zinc-finger.
 FT DOMAIN 519 590 MBD.
 FT DNA_BIND 622 634 A.T hook 1.
 FT DNA_BIND 643 655 A.T hook 2.
 FT DOMAIN 821 886 DDT.
 FT DNA_BIND 1159 1171 A.T hook 3.
 FT DNA_BIND 1377 1389 A.T hook 4.
 FT ZN_FING 1649 1699 PHD-type.
 FT DOMAIN 1783 1853 Bromodomain.
 FT DOMAIN 633 772 Lys-rich.
 FT DOMAIN 666 765 Coiled coil (Potential).
 FT DOMAIN 1185 1250 Glu-rich.
 FT DOMAIN 1263 1384 Pro-rich.
 FT DOMAIN 1732 1735 Poly-Arg.
 FT CONFLICT 574 574 L -> V (in Ref. 2).
 FT CONFLICT 700 700 L -> Q (in Ref. 2 and 3).
 FT CONFLICT 720 720 H -> Q (in Ref. 2 and 3).
 FT CONFLICT 727 738 SKAEKGTQK -> KKKKKKKKKK (in Ref. 2).
 FT CONFLICT 951 951 K -> R (in Ref. 3).
 FT CONFLICT 1005 1006 GR -> EG (in Ref. 3).
 FT CONFLICT 1163 1163 G -> S (in Ref. 1).
 FT CONFLICT 1166 1166 R -> L (in Ref. 1).
 FT CONFLICT 1172 1172 S -> F (in Ref. 1).
 FT CONFLICT 1178 1178 L -> F (in Ref. 1).
 FT CONFLICT 1202 1202 A -> V (in Ref. 1).
 FT CONFLICT 1292 1292 P -> L (in Ref. 1).
 FT CONFLICT 1295 1295 L -> F (in Ref. 1).
 FT CONFLICT 1313 1313 P -> L (in Ref. 1).
 FT CONFLICT 1407 1410 Missing (in Ref. 4).
 FT CONFLICT 1416 1416 R -> P (in Ref. 1).
 FT CONFLICT 1541 1541 R -> P (in Ref. 1).
 FT CONFLICT 1571 1571 E -> K (in Ref. 1).
 FT CONFLICT 1616 1616 V -> I (in Ref. 1).
 FT CONFLICT 1622 1622 E -> Q (in Ref. 1).
 FT CONFLICT 1629 1629 Q -> H (in Ref. 1).
 FT CONFLICT 1636 1636 Q -> H (in Ref. 1).
 FT CONFLICT 1739 1739 R -> K (in Ref. 1).
 FT CONFLICT 1754 1754 G -> R (in Ref. 1).
 SQ SEQUENCE 1878 AA; 208480 MW; 400970CA68234317 CRC64;

Query Match 98.9%; Score 9716; DB 1; Length 1878;
 Best Local Similarity 99.8%; Pred. No. 0;
 Matches 1856; Conservative 4; Mismatches 18; Indels 0; Gaps 0;

Qy 1 MEMEANEANDHFNFTGLPPAPASGLKPSGSSGGLYTNGPSMMNFPQGGKSLNGDVNVNG 60
 Db 1 MEMEANEANDHFNFTGLPPAPASGLKPSGSSGGLYTNGPSMMNFPQGGKSLNGDVNVNG 60
 Qy 61 LSTVSHTTTSGILNSAPHSSSTSHLHHPVSAYDCLWNYSQYPSANPGSNLKDPELLSQFS 120
 Db 61 LSTVSHTTTSGILNSAPHSSSTSHLHHPVSAYDCLWNYSQYPSANPGSNLKDPELLSQFS 120
 Qy 121 GGQYPLNGILGGSRQSPSSPHNTNLRAGSKQFWANGTHSPMGLNFDQSLEYDSPDQNF 180
 Db 121 GGQYPLNGILGGSRQSPSSPHNTNLRAGSKQFWANGTHSPMGLNFDQSLEYDSPDQNF 180
 Qy 181 EVCSGIHPDEAAEKEMTSSVVAENGTGLVCSLELEEXQPELKMCGYNGSVPSVSLHQEVS 240

Db 181 EVCSGHDHDEAAEKEMTSVVAENGGLVCSLEBEEQPELKCNGYGVSPVSESLHQVS 240
Qy 241 VLVPDPTVSCLDPPSHLPDQLEDTPILSEDLSLEPNSLAPFVSGGLYIGIDDTELMAEAD 300
Db 241 VLVPDPTVSCLDPPSHLPDQLEDTPILSEDLSLEPNSLAPFVSGGLYIGIDDTELMAEAD 300
Qy 301 KLPLXDSPIVIALDCPSLNNATAFSLADDSQTSISIFASPTSPVPLVGSVLQDNSFDLN 360
Db 301 KLPLXDSPIVIALDCPSLNNATAFSLADDSQTSISIFASPTSPVPLVGSVLQDNSFDLN 360
Qy 361 NGSDAEQEMETQSSDFPSSLTOPADQSSSTIQLHPATSPAVSPVTSFAVSLVSPASP 420
Db 361 NGSDAEQEMETQSSDFPSSLTOPADQSSSTIQLHPATSPAVSPVTSFAVSLVSPASP 420
Qy 421 EISPEVCPAASTVSPVSPVSPASSAVLPAVSLVLEPLTASVTSFKASPVTSFAVSPASP 480
Db 421 EISPEVCPAASTVSPVSPVSPASSAVLPAVSLVLEPLTASVTSFKASPVTSFAVSPASP 480
Qy 481 ASPANKOVSSFLETTADVEEITGEGLTASGSDVMRRRIATPEEVLRLPLQHGWRREVRK 540
Db 481 ASPANKOVSSFLETTADVEEITGEGLTASGSDVMRRRIATPEEVLRLPLQHGWRREVRK 540
Qy 541 KGSHRWQETWYGGCGKMKQFPEVIKYLSENLVHSVRRHFSFSPRMPVGDPEERDT 600
Db 541 KGSHRWQETWYGGCGKMKQFPEVIKYLSENLVHSVRRHFSFSPRMPVGDPEERDT 600
Qy 601 PEGLOWVLSAEIIPSRIOAITGKGRPRNTEKATKEVPKVRGRGRPPKVKITELLANK 660
Db 601 PEGLOWVLSAEIIPSRIOAITGKGRPRNTEKATKEVPKVRGRGRPPKVKITELLANK 660
Qy 661 TDNRPLKLEAQETLNEEDKAKIAKSKKMRQKQVORGECLTTIQOARNKRKQETKSLKH 720
Db 661 TDNRPLKLEAQETLNEEDKAKIAKSKKMRQKQVORGECLTTIQOARNKRKQETKSLKH 720
Qy 721 KEAKKSAEKEGKTQKLEKEKVKREKKEKVKMEKEEVTAKPACAKDKTLATQRRLL 780
Db 721 KEAKKSAEKEGKTQKLEKEKVKREKKEKVKMEKEEVTAKPACAKDKTLATQRRLL 780
Qy 781 EERQKQMILEMKPTEDMCLTDHQLPDRSRVPLGLTPSGAFSDCLTIIVEFLHSFGKV 840
Db 781 EERQKQMILEMKPTEDMCLTDHQLPDRSRVPLGLTPSGAFSDCLTIIVEFLHSFGKV 840
Qy 841 LGFDPKDVPSLVLQEGLLCOGDSLGEVQDILLVRLKAAALHDPGFPSYCQSLKILGEKV 900
Db 841 LGFDPKDVPSLVLQEGLLCOGDSLGEVQDILLVRLKAAALHDPGFPSYCQSLKILGEKV 900
Qy 901 SEIPLTRDNVSEILRCFLMAYGVXPALCDRLRTQFQAQPOOQKAAVLAFPVHELNGSTL 960
Db 901 SEIPLTRDNVSEILRCFLMAYGVXPALCDRLRTQFQAQPOOQKAAVLAFPVHELNGSTL 960
Qy 961 IINEIDKTESMSSYRKNKWIIVEGRRLRLKTVLAKRTGRSEVEMGRPEECLGRRRSSRIM 1020
Db 961 IINEIDKTESMSSYRKNKWIIVEGRRLRLKTVLAKRTGRSEVEMGRPEECLGRRRSSRIM 1020
Qy 1021 EETSGMBEEEEESIAAVPGRGRDRDGEVDATASSIPELERQIEKLSKRQFLFRKKLLHS 1080
Db 1021 EETSGMBEEEEESIAAVPGRGRDRDGEVDATASSIPELERQIEKLSKRQFLFRKKLLHS 1080
Qy 1081 SOMLRAVSLGQDRYRRRTWVLPYLAGI FVEGTEGNLVPPEVIKKTEDSLKVAHAASLNPA 1140
Db 1081 SOMLRAVSLGQDRYRRRTWVLPYLAGI FVEGTEGNLVPPEVIKKTEDSLKVAHAASLNPA 1140
Qy 1141 LPSMKWELAGNNTTASSPARASRLKTKPGMQPRHFKPSVPRGDSQPQQAQLQPEAQL 1200
Db 1141 LPSMKWELAGNNTTASSPARAGRPDKTKPGMQPRHLKSPVRGDSQPQQAQLQPEAQL 1200
Qy 1201 HVPAPQPOLQLQASHKGFLEQEGSPLSQSQHDLQSQAFSLWSLQTSQSHSSLSLSSV 1260
Db 1201 HVPAPQPOLQLQASHKGFLEQEGSPLSQSQHDLQSQAFSLWSLQTSQSHSSLSLSSV 1260
Qy 1261 LTPDSSPGKLDPAQSPPEEPDEAESSPDLQAFWFNISQMPCNAAPTPPPLAYSEDPQ 1320
Db 1261 LTPDSSPGKLDPAQSPPEEPDEAESSPDLQAFWFNISQMPCNAAPTPPPLAYSEDPQ 1320

Db 1261 LTPDSSPGKLDPAQSPPEEPDEAESSPDLQAFWFNISQMPCNAAPTPPPLAYSEDPQ 1320
Qy 1321 TPSQQOLASSKPMNRPSAANPCSPVQFSSTPLAGLAKRRRAGDPCGEMPSQSTGLGQPKRR 1380
Db 1321 TPSQQOLASSKPMNRPSAANPCSPVQFSSTPLAGLAKRRRAGDPCGEMPSQSTGLGQPKRR 1380
Qy 1381 GRPSPKFFKQMEQRYLTQLTAQVPVPEMCSQWMIIDPEMLDAMLKALHPRGIREKALHK 1440
Db 1381 GRPSPKFFKQMEQRYLTQLTAQVPVPEMCSQWMIIDPEMLDAMLKALHPRGIREKALHK 1440
Qy 1441 HLNKRRDPLQSVCLURPSADPIFEPRQLPAFOEGIMSNKPEKTYETDVLAVLQWVEELQR 1500
Db 1441 HLNKRRDPLQSVCLURPSADPIFEPRQLPAFOEGIMSNKPEKTYETDVLAVLQWVEELQR 1500
Qy 1501 VIMSDLOIRGWTCFSPDSTREDLAYCEHLSQSDQEDITWRGFGREGGLAPQRTTNPDLAV 1560
Db 1501 VIMSDLOIRGWTCFSPDSTREDLAYCEHLSQSDQEDITWRGFGREGGLAPQRTTNPDLAV 1560
Qy 1561 MRLAALQNVKRRYLREPLWPTHEVLEKALLSTPNGAPEGTTTTEISYEITPRIRWROT 1620
Db 1561 MRLAALQNVKRRYLREPLWPTHEVLEKALLSTPNGAPEGTTTTEISYEITPRIRWROT 1620
Qy 1621 LORCRSAHVCLCLGHLERSIAWEKSNKVTCLVCRKGDNDDEFLLLCDCGDRGCHYCHR 1680
Db 1621 LORCRSAHVCLCLGHLERSIAWEKSNKVTCLVCRKGDNDDEFLLLCDCGDRGCHYCHR 1680
Qy 1681 PKMEAVPEGDFWCTVCLAQVQVEGFTQKPGPKGQKRSKGSYLNFSFGDGRRRVLLKG 1740
Db 1681 PKMEAVPEGDFWCTVCLAQVQVEGFTQKPGPKGQKRSKGSYLNFSFGDGRRRVLLKG 1740
Qy 1741 RESPAAGPRYSEERLSPSKRRRLSNRNHSDLTFCCEILMEMESHDAAWPFEPVNPRLV 1800
Db 1741 RESPAAGPRYSEERLSPSKRRRLSNRNHSDLTFCCEILMEMESHDAAWPFEPVNPRLV 1800
Qy 1801 SGYRIIKNPMDFTMRERLLRGYTSSEEPAAADALLVDFNCOTFNEDDSEVGKAGHLMR 1860
Db 1801 SGYRIIKNPMDFTMRERLLRGYTSSEEPAAADALLVDFNCOTFNEDDSEVGKAGHLMR 1860
Qy 1861 RFFESRWEFFYQKQANL 1878
Db 1861 RFFESRWEFFYQKQANL 1878
ID Q68DI8 PRELIMINARY; PRT; 1905 AA.
AC Q68DI8;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DE Hypothetical protein DKFp781B109.
GN Name=DKFp781B109;
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=Fetal kidney;
RG The German cDNA Consortium;
RA Bloecher H., Boescher M., Brandt P., Mewes H.W., Weil B., Amid C.,
RA Osanger A., Fobo G., Han M., Wiemann S.;
RL Submitted (AUG-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; CR749379; CAH18232.1; -;
DR InterPro; IPR000637; A-T_hook.
DR InterPro; IPR001487; Bromodomain.
DR InterPro; IPR004022; DDT.
DR InterPro; IPR011011; FYVE_PHD_Znf.
DR InterPro; IPR001739; Methyl-CpG_bind.
DR InterPro; IPR001965; Znf_PHD.
DR Pfam; PF02178; AT_hook; 4.
DR Pfam; PF00439; Bromodomain; 1.
DR Pfam; PF02791; DDT; 1.

RESULT 3
BAZA_MOUSE
ID BAZA_MOUSE STANDARD; PRT; 1850 AA.
AC Q91VE5;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Bromodomain adjacent to zinc finger domain 2A (Transcription
termination factor-1 interacting protein 5) (TFP-1 interacting protein
5) (Tip5).
DE (Tip5).
CN Name=Baz2a; Synonyms=Tip5;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21423468; PubMed=11532953; DOI=10.1093/emboj/20.17.4892;
RA Strohnner R., Nemeth A., Jansa P., Hoffmann-Rohrer U., Santoro R.,
RA Laengst G., Grummt I.;
RT "NORC-a novel member of mammalian ISWI-containing chromatin remodeling
machines".
RL EMBO J. 20:4892-4900(2001).
CC -!- FUNCTION: May play a role in transcriptional regulation
interacting with ISWI. May serve a specific role in maintaining or
altering the chromatin structure of the rDNA locus.
CC -!- SUBUNIT: Together with ISWI/SNF2h, it forms a complex termed NORC
(nucleolar remodeling complex).
CC -!- SUBCELLULAR LOCATION: Nuclear. Colocalizes with the basal RNA
polymerase I transcription factor UBF in the nucleolus.
CC -!- SIMILARITY: Belongs to the WAF family.
CC -!- SIMILARITY: Contains 4 A-T hook DNA-binding repeats.
CC -!- SIMILARITY: Contains 1 bromodomain.
CC -!- SIMILARITY: Contains 1 DDT domain.
CC -!- SIMILARITY: Contains 1 methyl-CpG-binding (MBD) domain.
CC -!- SIMILARITY: Contains 1 PHD-type zinc finger.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb.ch/announce/>
or send an email to license@isb.ch)

DR EMBL; AJ309544; CAC59992.1; -;
DR HSP; Q91G0; 1F62.
DR MGD; MGI:2151152; Baz2a.
DR GO; GO:0005731; C:nucleolus organizer complex; ISS.
DR GO; GO:0003677; F:DNA binding; ISS.
DR GO; GO:0030528; F:transcription regulator activity; ISS.
DR GO; GO:0006338; P:chromatin remodeling; ISS.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; ISS.
DR InterPro; IPR000637; A-T hook.
DR InterPro; IPR001487; Bromodomain.
DR InterPro; IPR004022; DDT.
DR InterPro; IPR001369; Methyl-CpG_bind.
DR InterPro; IPR001965; Znf PHD.
DR Pfam; PF02178; A-T hook; 4.
DR Pfam; PF00439; BrComodomain; 1.
DR Pfam; PF02791; DDT; 1.
DR Pfam; PF01429; MBD; 1.
DR Pfam; PF06628; PHD; 1.
DR PRINTS; PR00503; BROMODOMAIN.
DR SMART; SM00384; AT hook; 4.
DR SMART; SM00297; BROMO; 1.
DR SMART; SM00571; DDT; 1.
DR SMART; SM00391; MBD; 1.
DR SMART; SM00249; PHD; 1.
DR PROSITE; PS00633; BROMODOMAIN_1; FALSE_NEG.
DR PROSITE; PS50014; BROMODOMAIN_2; 1.
DR PROSITE; PS50827; DDT; 1.

DR PROSITE; PS50982; MBD; 1.
DR PROSITE; PS01359; ZF PHD 1; FALSE_NEG.
DR PROSITE; PS50016; ZF PHD 2; 1.
KW Bromodomain; Coiled Coil; DNA-binding; Nuclear protein; Repeat;
KW Transcription regulation; Zinc-finger.
FT DOMAIN 499 570 MBD.
FT DNA_BIND 602 614 A.T hook 1.
FT DNA_BIND 623 635 A.T hook 2.
FT DOMAIN 800 865 DDT.
FT DNA_BIND 1137 1149 A.T hook 3.
FT DNA_BIND 1351 1363 A.T hook 4.
FT ZN_FING 1823 1873 PHD-type.
FT DOMAIN 1755 1825 Bromodomain.
FT DOMAIN 613 738 Lys-rich.
FT DOMAIN 647 774 Coiled coil (Potential).
FT DOMAIN 1006 1012 Poly-Glu.
SQ SEQUENCE 1850 AA; 205597 MW; E5DD4FED1D46DAE1 CRC64;

Query Match 79.3%; Score 7795; DB 1; Length 1850;
Best Local Similarity 80.4%; Pred. No. 6.7e-313;
Matches 1503; Conservative 111; Mismatches 202; Indels 54; Gaps 9;

QY 43 MNFPQCKSLGNDVNVNGLSTVSHHTTSGILNSAPHSSSTSHLHHPVAVYDCLWNYSP 102
DB 1 MSLPQCKSLGNDVNVNGLSTVSHHTTSGILNSAPHSSSTSHLHHPVAVYDCLWNYSP 60

QY 103 SANPGSNLKDPLLQSFGQYPLNGILGGRQSPSPSHNTNLRAGSKFWANGTHSPWG 162
DB 61 SANFGNLLKDPLLQSFGQYPLNGILGGRQSPSPSHNTNLRAGSEFWANGTQSPWG 120

QY 163 LNFDSQELYSFPPQNFEE-----EVCSGIHPDRAA 192
DB 121 LNFDSQELYSFPPQNFEEVMPNGPPSPPTSPQTSPLMGSSIQTAPSQDVSDIHDEAA 180

QY 193 EKEMTSVVAENGTLGVCSLELEEXQPELKMCGYNGSVPSVSLHQEVSLVPDPTVSCLD 252
DB 181 EKELTSVVAENGTLGVCSLELEEXQPELKMCGYNGSVSVSLHQEVSLVPDPTVSCLD 240

QY 253 DPSHLPQLEDTPLTSLSDSLPFPNSLAPEPVSGLYGIDDTLMGAEDKLPLKDSPLISA 312
DB 241 DPSHLPQLEDTPLTSLSDSLPFPNSLAPEPVSGLYGIDDTLMGAEDKLPLKDSPLISA 300

QY 313 LDCPSLNNAATFSLADDSTQTSSTISFASPTSPFVLGESVLDNSFDLNGSDAEQEMET 372
DB 301 LDCPALSNANAFSLADDSTQTSSTISFVSPPTSPFVLGESVLDNSFDLNGSDAEQEEIET 360

QY 373 QSSDPPPSLTOPADQDSTTIOLHPATSPAVSPPTSPAVSLVVSPPASPEISPEVCPAAST 432
DB 361 QSSNFQRLTEPAPDQPPSTQLH-----PAVSPTASPAAL-----TASAEISPAVSPVASS 412

QY 433 VVSPAVSVVSPASSAVLPAVSLVPLTASVTSKASPVTSPPAAAFPTASPAKDVSSFL 492
DB 413 PVPPEVFAVSPASSPALPAISLEASMTTPVTSFGGSEPPSPAAAFQTVSPARKNVSSAP 472

QY 493 ETTADVBEITGEGLTASGSDVMRRRIATPEVRLPLQHGWRREVRIRKKGSHRWQETWY 552
DB 473 KARADRETTGGAVAVSGSDVLRATPEVRLPLQHGWRREVRIRKKGSHRWQETWY 532

QY 553 YGPGCKRMKQPEVVKYLSRNLVHSVRREHFSFSPMPVGVDFEEDTPPEGLQWVLSAE 612
DB 533 YGPGCKRMKQPEVVKYLSRNLVHSVRREHFSFSPMPVGVDFEEDTPPEGLQWVLSAE 592

QY 613 EIPSRIOAITCKGRPRNTEKAKTKEVPKVRGRGRPPKVKITELINKTDNRPLKLEAQ 672
DB 593 EIPSRIOAITCKGRPRNNEKAKNKEVPKVRGRGRPPKIKPELLNKTDNRPLKLEAQ 652

QY 673 ETLNEEDKAKIATKCKKQKQVQGECLTTTQGOARKNRKQETKSLKHKEAKKSKAEKE 732
DB 653 EILSEDDKAKTKKQKQKQVQGESQTPVQGOARKNRKQDTKSLKQDKTKKLKAEKE 712

QY 733 KGKTKQEKLEKVKREKKEKVKMKKEBVTYKAPKADKTLATQRRLEERKQKQMILEE 792
DB 713 KWKTKQEKLEKVKREKKEKVKAKGEG-PRARPSCRADKTLATQKRLLEEQRQQAILEE 771

```
Qy 793 MKKPTEDMCLTDHQPLPDSRVPGTLPSGAFSDCLTIVEFLHSGKVLGFDPAKDVPSL 852
Db      |||||
Qy 772 MKKPTEGNCLSDHQPLPDTTRIPGLTSSRAFSDCLTIVEFLHSGKVLGFDLTKDVPSL 831
Db      |||||
Qy 853 GVLQEGLLCOGDSLGCEVDLLVRLLLKAALHDPGPPSYCQSLKILGEKVESEIPLTRDNVSE 912
Db      |||||
Qy 832 GVLQEGLLCOGDSLDKVDQDLLVRLLLKAALHDPGLPPYQCQLKIIGKMKSEIPLTRDNVSE 891
Db      |||||
Qy 913 ILRCFLMAYGVKPCALCDRLRTPQAOQPOQAAPVLAAPVHELNGSTLIINEIDKLTESM 972
Db      |||||
Qy 892 ILRCFLMAYRVEPPFCDSRLRTPQAOQPOQAAILAFLVHELNSTIINEIDKLTESV 951
Db      |||||
Qy 973 SSYRKNKWIYBGRLLRLTLAKTRGSEVEMGRPEECLGRRRSRRIEETSGMEEBEE 1032
Db      |||||
Qy 952 SSCRNKWIYBGRLLRLTALAKTRGPEVMEGAEDGLGRRRSRIEETSGIEEBEE 1011
Db      |||||
Qy 1033 ESIAVPGRRGRGGEVDATASSIPELROQIEKLSKROLFRKKLLHSSQMLRAVSLGQD 1092
Db      |||||
Qy 1012 ENTAVHGRRGRKEGEIDVAASSIPELERHIEKLSKROLFRKKLLHSSQMLRAVSLGQD 1071
Db      |||||
Qy 1093 YRYYRWLPLVLAGTFVEGTGNLVPEVVIKKTDSLKVAHAASLNPAFLPMKVELAGSN 1152
Db      |||||
Qy 1072 YRYYRWLPLVLAGTFVEGSGSTVTEDEIRKQETESLMEVVTSTPSSARASVKRELTCGN 1131
Db      |||||
Qy 1153 TTASPARARSRPLKTKPGFQWPRHFKSPVRGQDSEQPOAQLHVPQAQPQPOLQ 1212
Db      |||||
Qy 1132 -ASTSPARSRRPRKPKGCSLQPOHLQSTIRECSEQAQTQVPE-----POPQLQA 1182
Db      |||||
Qy 1213 QLQSH----KGLFQEGSPLSGOSQHDLSQSAFLSWLSQTSQSHSLSSSVLTPDSSPG 1268
Db      |||||
Qy 1183 PTQHLQPSGFLRPEGSPFLSGOSQHDLSQSAFLSWLSQTSQSHSLSSSVLTPDSSPG 1242
Db      |||||
Qy 1269 KLDPAPOPPPEPDEBESPDLOAFWPNLSAOMPNAATPPPLAVSEDOPTSPQOLA 1328
Db      |||||
Qy 1243 KLDAPSOSLEPEPDEAQCSPGPGPWFNFSAQPCDAAPTPPPAVSEDOPTPSQLLA 1302
Db      |||||
Qy 1329 SSKPMNRPSAANPCSPVQSTPLAGLAPKRAGDPGEMPOSPTGLGQPKRRGPPSPKFF 1388
Db      |||||
Qy 1303 SSKPMNTPGAANPCSPVLSSTHLPFGTTPKRLSGDSEMSQPTGLGQPKRRGPPSPKFF 1362
Db      |||||
Qy 1389 KQEQRYLTQLTQAQVPPMCSGWWWIPDPBMLDAMLKALHPRGIKREKALHKLHNRDF 1448
Db      |||||
Qy 1363 KQVEQHYLTQLTQAQIPPEMCSGWWWIRDPETDLVLLKALHPRGIKREKALHKLHKKHF 1422
Db      |||||
Qy 1449 LQEVCLRPSADPIPEPQPLPAFQEGINSWSPEKTYETDVLAVQWVELEORVIMSLOI 1508
Db      |||||
Qy 1423 LQEVCLQPLTDPIPEPNELPALGEGVMSWSPEKTYETDVLAVQWVELEORVIMSLOI 1482
Db      |||||
Qy 1509 RGWTCPSDSTREDLAYCEHLSDSQEDITWRGPGREGLAPOKKTNPDLAVMLAALQ 1568
Db      |||||
Qy 1483 RGWTCPTDSTREDITYCEHLPDSGEDIPWGRGREGTVPORQNNPDLAVMLAVLEQ 1542
Db      |||||
Qy 1569 NVKRYRLREPWPHEVVLEKALLSTNGAPEGTTTETISYITPRIRIWRQTLQRCRSAA 1628
Db      |||||
Qy 1543 NVRRYRLREPWAAHEVVVEKALLSTNGAPDGTSTETISYITPRVVRWQTLERCSSAA 1602
Db      |||||
Qy 1629 HVCICLGHLSIAWEKSVNKTCLVCKGDNDFLLCDGCDRGCHYIHRPKMEAVPE 1688
Db      |||||
Qy 1603 QVCLCMGQLERSIAWEKSVNKTCLVCKGDNDFLLCDGCDRGCHYIHRPKMEAVPE 1662
Db      |||||
Qy 1689 GDMCTVCLAQOQVEFTQKGFPRKGRKSGYSLNPFSEGDGRRRRLVLLKGRSPAAGP 1748
Db      |||||
Qy 1663 GDMCAVCLSQOQVEEYTORPGFKRQGRKSSPPLTFPEGDS-RRRMLSRSDSPAV-P 1720
Db      |||||
Qy 1749 RYSERLSPSKRRRLSMNHSDLTFCEIILMEMESHDAAPFPXEPVNPRLVSGVRRIK 1808
Db      |||||
Qy 1721 RYPEDGLSPKRRRSMHSHSDLTFCEIILMEMESHDAAPFPVNPRLVSGVRRIVK 1780
Db      |||||
Qy 1809 NPMDFSTMRKLLRGYTTSSBEFAADALLVFDNCQTFNEDDSYGVKAGHIMRRFFESRWE 1868
Db      |||||
Qy 1781 NPMDFSTMRKLLRGYTTSSBEFAADALLVFDNCQTFNEDDSYGVKAGHVMRRFFESRWE 1840
Db      |||||
```

```
Qy 1869 EFYQKGQOANL 1878
Db      |||||
Qy 1841 EFYQKGQOANL 1850
Db      |||||

RESULT 4
Qy Q6Y194 PRELIMINARY; PRT; 1698 AA.
Ac Q6Y194;
Dt 05-JUL-2004 (TrEMBLrel. 27, Created)
Dt 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
Dt 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
De Putative chromatin remodelling factor BAZ2A.
Os Xenopus laevis (African clawed frog).
Oc Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Oc Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
Oc Xenopodinae; Xenopus.
Ox NCBI_TaxID=8355;
Rn [1]
Rp SEQUENCE FROM N.A.
Ra Ruzov A., Meehan R.;
Rl Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
Dr EMBL; AY145834; AAN61105.1; -.
Dr GO; GO:0005634; C:nucleus; IEA.
Dr GO; GO:0003677; F:DNA binding; IEA.
Dr GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
Dr InterPro; IPR00637; A-T hook.
Dr InterPro; IPR001487; Bromodomain.
Dr InterPro; IPR004022; DDT.
Dr InterPro; IPR001739; Methyl-CpG_bind.
Dr InterPro; IPR001965; ZnF_PHD.
Dr Pfam; PF02178; AT hook; 3.
Dr Pfam; PF00439; Bromodomain; 1.
Dr Pfam; PF02791; DDT; 1.
Dr Pfam; PF01429; MBD; 1.
Dr Pfam; PF00628; PHD; 1.
Dr PRINTS; PR00503; BROMODOMAIN.
Dr SMART; SM00384; AT hook; 3.
Dr SMART; SM00297; BROMO; 1.
Dr SMART; SM00571; DDT; 1.
Dr SMART; SM00391; MBD; 1.
Dr SMART; SM00249; PHD; 1.
Dr PROSITE; PS50014; BROMODOMAIN_2; 1.
Dr PROSITE; PS50827; DDT; 1.
Dr PROSITE; PS50016; ZF_PHD_2; 1.
Dr PROSITE; PS50016; ZF_PHD_2; 1.
Sq SEQUENCE 1698 AA; 192135 MW; F62D613AD9611B6F CRC64;

Query Match 38.4%; Score 3777; DB 2; Length 1698;
Best Local Similarity 44.9%; Pred. No. 2.3e-147;
Matches 874; Conservative 254; Mismatches 482; Indels 336; Gaps 56;

Qy 7 BANDHFNFTGLPPAPAASGLKPSGEGLYTNGSPMFPQOGKSLGNDVNVNGLSTVSH 66
Db      |||||
Qy 2 ETQNHLSCTGTPLTNASGLKLLPQSVEPYVTNTLTLSFSQHGKVLNLDLPVNGLSIPS- 60
Db      |||||
Qy 67 TTTSGLINSAPHS--STSHLHPSVA--YDCLWNYSQYPSANPGSL-KDPPLLSQF-- 119
Db      |||||
Qy 61 -----LSNNAHGTLISNEHITHNSASNNFDYLWK-----KPKSVHKDHSVSNQFLV 107
Db      |||||
Qy 120 SGGOYPL-NG-----ILGG--SRQPS--PSHN-----TNLRAGSQKFWANGTH 158
Db      |||||
Qy 108 NGSTCPTNGPSIKRSQETILGNCTSNQTASFNVSSHSCVSPNEANYKMVTA---ANGAH 164
Db      |||||
Qy 159 SPMGLNFDSELDYDPPQNFEEVCSGHIHPDEAAEKMTSVVAENGTV-----LV 208
Db      |||||
Qy 165 ----YGFYEASLPVSVSASNNQSV-SG-----DNLVKERT--VEENEAGATEERNHTPELA 213
Db      |||||
Qy 209 CSLSELEXQPELKMCGYNGSV--PSVESLHQVSVLPDPTVSCLDPSHLPDQLEDTPPI 266
Db      |||||
Qy 214 EALEPSTQTDLNICKEYKSVLDPIQOS---TNPLLP-PDVSNLDDPSQLPSQLGSHL 269
Db      |||||
Qy 267 LSEDSLEPFPNS-LAPEPVSGGLYCIDDTLWLGAEADKLPLXDSPVISALDCPFLNNATAFS 325
Db      |||||
```

Db 270 LNEDELEPFADLIQDPISSMYDLEEEVSGTKQKDPQDIT---SSLDCTYSLSAAPP 326
Qy 326 LLAADSQTSITSIFASPTSPVGLGESVLQDNFGLNMGSDABQEEMETQSSDFPPLSTQPA 385
Db 327 LVAEDTNDSSVLFNASSASPVLDGSMQESASEM--GEDPEGSKAERPVSQ--PENVSQ-- 381
Qy 386 PQDSSTIQHLPATSPAVSPPTSPAVSLVVSAPASPEISPEVCPAASSTVVSAPVSVSPA 445
Db 382 --DEMTIE-----NTSPSPCVAAE----- 398
Qy 446 SSAVLPAVSLVPLTASVTPSPKASPVTPSPAAAPTASPAKNDVSFLETTADVEEI--TGE 504
Db 399 -----HEEELEFGE 408
Qy 505 GLTAGSGDVMRRRIATPEEVLRLPLQHGWRREVRIKKGSHRWQGETWYFGCGKRMKQFP 564
Db 409 -----VKGTISRRIATPEQVCFPLQHGWRREVRIKKGSHRWQGETWYFAPCGKRMKQFP 463
Qy 565 EVIKYLSNLVHSVREHFSPSPMPVCDPFEERTPEGLQWVQLSABEISPSRTOAITGK 624
Db 464 EVIKYLSNAGPFPVREHFSPSPMPVCDPFEERTPEGLQWVQLSABEISPSRTOAITGK 523
Qy 625 RGRPNTKATKEVPKVRGRGRPPVKVITELNKTQNRPLKLEAOETLNEEDKAKIA 684
Db 524 RGRPNLEKAKAKE--OKAKRGRGRPPVKVIMDLISKADAKLRLKLENDIILSDSEKVLQ 582
Qy 685 KSKKQVRQKQVORGECLTTIOGAQRNKRQETKSLKHKEAKKSKAEKKGTKQEKLEK 744
Db 593 KKKQVRK-----ARQEALEAAKKLEKEK-----EEKKQIKQAKQNEK 626
Qy 745 VIREKKEKVKKEEVTYKAPKAD--KTATQRLLEERQKQKQMIEMKKPTEDMCLT 803
Db 627 AKNOEKKTRQPKO---KAPVQKPDRLQQAQRLKRRKQOPILEELKKPTEDMCLP 683
Qy 804 DHOPLPDFSRVPLGLPSPGASDCLTIVFELHSGKVLGFGPAKDVPSLVLOGLLQCG 863
Db 684 DHOQLPDFPCVPLGLPSPGASDCLTIVFELHSGKVLGFGPAKDVPSLVLOGLLQCG 743
Qy 864 DSLGEVDLLVRLKAAHLDFGFPSPYCOSLKILGEKVSEIPLTRDNVSEIILRCFLMAYGV 923
Db 744 DSLGEVDLLVKLLQAAINFGLPYQCQSLKILGEKVSEIPLTRDNVSEIILRCFLMAYGV 803
Qy 924 XPALCDRLRTPFOQAOPQQAALVAPVPHLNGSTLIINEIDKTLSESSSYRKNKNWIVE 983
Db 804 DIQLCDSLRSHFPQAHAPHIKTALIAFTVNLNASSLIISBIDKTLNNSHYRKHKWIIE 863
Qy 984 GLRLKTVLAKTRGSEVEMGRPECLGRSSRIMEETSGMEHEEESIAAVPGRG 1043
Db 864 GKIRRLKFKALSKRSSESSQITTTTEVSL--RRSER-----NABENDELSIDSSAIQKO 916
Qy 1044 RRDGEVDATAS--SIPELEROIEKLSKROLFFRKLLHSSQMLRAVSLGQDRYRRYVLP 1102
Db 917 YVQEVDPDPTSVVLELROIEKLTQKQMFPRKILGSSQRLRTVCLGQYRRCYRWMLP 976
Qy 1103 YLAGIFVGE-----TE-----GNLVPEVITKKTETDSLKVAHAHSLNPAFLSMKMWELAGS 1151
Db 977 HLGIGIFIEGLPESABPTBEAALNDIEASVKTEDKSPGLCKTSGHP-----R 1025
Qy 1152 NTTASSPARASRLPKTKPGWQPHFKSPVRGQDSEQPOQLQPEAQHVPAPQPOLO 1211
Db 1026 NSTA-----EPEQNSTSCHSDSKDEP----- 1048
Qy 1212 LQLQSHKGLFEGQSPSLGOSQHDLSQAFSLMSLQTSQSHSLSSLLSSVLPDPSGPKLD 1271
Db 1049 -----SGVTQDFPNSVPLTNQDLSQSVFLSWL--TKNQTSIMDSVILPESPPHSE 1100
Qy 1272 PAP-----SQPPEPEDEAESPDLOAFWNFISAQMPCN-----AAPTPLA 1314
Db 1101 STPIISSEATEKPEQ-----MLPIISRTPCRNHNQGLSTHSSNLSPPSPATA 1148
Qy 1315 VSEDOP-----TSPQOLASSKPMWRPSAANPCSPVQFSSTPLAGLAPKRRAGDPGEMOSP 1371
Db 1149 TSVKQVNEFTBEAQTFATSLPSN--STPCHVCVNSGKSGSTASHEITLTSNIIHLHSE----- 1202

Qy 1372 TGLQOPKXRRGPPSPKFKQMEQRYLTQLTQAPVPEMCSGWWIIPDPBMLDAMLKALHPR 1431
Db 1203 ---KEKRRGRPPSKLLKQIEQKYNYQLIERPIIPAGVRQKWWIKDPAMLSLLKALHPR 1258
Qy 1432 GIREKALHKLHGRDFLOEVLCPSPADPIFEPQLPAF---QEGINSMSPEKTYETDL 1488
Db 1259 GIREKTLHKYJTKHLQHLKEMCARPASDALPKFTVPDGHVRSQETLDRWSVTDLTFFQVDL 1318
Qy 1489 AVLQWBELEQRVMSDLQIRGWTCPSPDSTREDLAYCEHLSDSQEDITMRGPGREGIAP 1548
Db 1319 SALQWVEDLEQRVMSDLQIRGWSFAPSVRTDLKYEHQLEADDDITVK--VKREDCL 1377
Qy 1549 QRKTTNPLDLAVMLAALAEQNVKRYLREPUPHETHEVLEKALLSTNGAPEGTTTEISY 1608
Db 1378 YRESSNPLDLAVLILCLEENVERKYLKEPLWLFSEVQHEKVVITNPEDPL--STTEIY 1435
Qy 1609 EITPRIRIMQTLORCSAAHVCILGHLERSIAWEKSVNKKVTCVLCRKGNDDEFLLCD 1668
Db 1436 SITSRLRLMQTVRCRSAQLSLQQLERSIAWERSLNKVTCLYCRKGNDDELLLCD 1495
Qy 1669 GCDRGCHYCHRPQWEAVPEGDWFCVCLAOQVEGEFTQKPGFKRGOKRKGSLNFS 1728
Db 1496 SCDRGCHYCHRPQWNEIPEGDFCPTCISLQSESEFLRSGSSKRIK----CTVRPTE 1551
Qy 1729 GDGR--RRVLLKGRESPAAGPRYSEERLSPSKRRRLSMRNNHSDLTFCFCEIILMEMESH 1786
Db 1552 DSPSKPSRR-----REHPTAS--QSPGESPAKCKRMGTRSQSPDLTFCFCEIILMELESHE 1605
Qy 1787 ANWPXEPVNPRLVSGYRRRIKKNPMDFTWRERLLRGYTSSEFPAADALLVFNOCFN 1846
Db 1606 DANWFLPEVNPRLVPGYRKIKKNPMDFTWRHKLNGYSCRCEFAEDAEILFNSCOLFN 1665
Qy 1847 EDDSEVGKAGHIMRRFPFESRWEFFYQ 1872
Db 1666 EDESVDGKAGLILKKFYDARWEEFSQ 1691
RESULT 5
Qy 080U42 PRELIMINARY; PRT; 886 AA.
AC Q80U42
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE MKTAA0314 protein (Fragment).
GN Name=Baz2a; Synonyms=mKIAA0314;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=2259291; PubMed=12693553;
RA Okazaki N., Kikuno R., Ohara R., Inamoto S., Aizawa H., Yuasa S.,
RA Nakajima D., Nagase T., Ohara O., Koga H.;
RT "Prediction of the coding sequences of mouse homologues of KIAA gene:
RT II. The complete nucleotide sequences of 400 mouse KIAA-homologous
RT cDNAs identified by screening of terminal sequences of cDNA clones
RT randomly sampled from size-fractionated libraries.";
RL DNA Res. 10:35-48(2003).
DR EMBL; AK122243; BAC65525.1; --
DR HSSP; O9UIG0; 1F62.
DR MGD; MGI:2151152; Baz2a.
DR GO; GO:0006338; P:chromatin remodeling; TAS.
DR InterPro; IPR000637; Art hook.
DR InterPro; IPR001487; Bromodomain.
DR InterPro; IPR011011; FYVE PHD_Znf.
DR InterPro; IPR001965; Znf_PHD.
DR Pfam; PF02178; AT_hook; 2.
DR Pfam; PF00439; Bromodomain; 1.
DR Pfam; PF00628; PHD; 1.

Db 182 PSFTSPQTSPLGSSIQTFAPSDVSDIHPDENAEEKELTSVAENGTGLVLSLELEE 241
 Qy 217 QPELKMCGYNGSVSPVESLHQEVSVLPDPPTVSCLDLDPDPSHLPDQLEDTPILSLESLEFPN 276
 Db 242 QPELKMCGYNGSVSPVESLHQEVSVLPDPPTVSCLDLDPDPSHLPDQLEDTPILSLESLEFPD 301
 Qy 277 SL-APEPVSGLYGIDDTFELMGAEDKPLXDSPVISALDCPSLNNATAFSLIADDSQTST 335
 Db 302 SLAAAEFVSGLYGIDDAELMGAEDKPLXDSPVISALDCPSLNNATAFSLIADDSQTSA 361
 Qy 336 SIFASPTSPVSGVLDQNSFDLNGSDAQEEMETQSSDFPPLSLTOPADQDQSTTQLH 395
 Db 362 SIFVSPSPVSGVLDQNSFGLNSCSDSEQEIEIETQSSNFQRPPLTEPAPDQPPSTQLH 421
 Qy 396 PATSPAVSPPTSPAVSLVSPAASPEISPEVCVPAASVTVSPAVSVSPASSAVLPAVSL 455
 Db 422 ----PAVSPTASPAASL-----TASAEISPAVSPVASSPVPEFVAVSPASSPALPAISL 473
 Qy 456 EVPLTASVTSKASPVTSAPAAFTASPAKDVSSFLETTADVEEITCEGLTASGSDVM 515
 Db 474 EASMTTPTVTSPOGSPSPAPAAFTQVSPAKDVSSAPKARADRETTGGAVAVSGSDVL 533
 Qy 516 RRRIATPEEVLPLQHGWRREVRIKKGSHRWQGETWYVPCGKMKQPEVIKYLRLV 575
 Db 534 KRRIATPEEVLPLQHGWRREVRIKKGSHRWQGETWYVPCGKMKQPEVIKYLRLV 593
 Qy 576 HSVREHFSFPRMPVGDFFERTDTEGLQVQLSAEIPSRIOAITKGRGRPRNTEKAK 635
 Db 594 HSVREHFSFPRMPVGDFFERTDTEGLQVQLSAEIPSRIOAITKGRGRPRNTEKAK 653
 Qy 636 TKEVPKVRGRGRPPKVKITELLNKTDNRPLKLEAQTNEEDKAKIAKSKKKMR 695
 Db 654 NKEVPKVRGRGRPPKVKITELLNKTDNRPLKLEAQTNEEDKAKIAKSKKKMR 713
 Qy 696 RGECLTTTQGOARNRKQETSLKHKEAKKXSAEKEGKTKQEKLEKVKREKKE 751
 Db 714 RGECLTTTQGOARNRKQETSLKHKEAKKXSAEKEGKTKQEKLEKVKREKKE 769

RESULT 7
 Q80VL8
 ID Q80VL8 PRELIMINARY; PRT: 709 AA.
 AC Q80VL8, Q96EA1; Q96SQ8; Q9P252; Q9Y4N8;
 DT 01-JUN-2003 (TREMBlrel. 24, Created)
 DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
 DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
 DE Hypothetical protein (Fragment).
 GN Name=Baz2a;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Eye;
 RA Strauberg R.;
 RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC048415; AAA48415.1; --
 DR MGD; MGI:2151152; Baz2a.
 DR GO; GO:0006338; P:chromatin remodeling; TAS.
 DR InterPro; IPR000637; A+T hook.
 DR InterPro; IPR001739; Methyl-CpG_bind.
 DR Pfam; PF02178; AT hook; 2.
 DR Pfam; PF01429; MBD; 1.
 DR SMART; SM00384; AT_hook; 2.
 DR SMART; SM00391; MBD; 1.
 KW Hypothetical protein.
 FT NON TER 709
 SQ SEQUENCE 709 AA; 76142 MW; 23B91AB778F92C4B CRC64;

Query Match 30.1%; Score 2960.5; DB 2; Length 709;
 Best Local Similarity 80.7%; Pred. NO. 3.4e-114;
 Matches 578; Conservative 32; Mismatches 67; Indels 39; Gaps 4;

Qy 7 EANDHFNFTGLPPAPAAAGLKPSPSSGEGLYTNGSPMNFPQOGKSLNGDVNVNGLSTVSH 66
 Db 2 EANDHFNFTGLPPAPAAAGLKPSPSSGEGLYTNGSPMNFPQOGKSLNGDVNVNGLSTVSH 61
 Qy 67 TTTSGIILNSAPHSSSTSHLHHPSVAVDCLWNYSOYPSANPGSNLKDPLLSQFSGGOYPL 126
 Db 62 TTTSGIILNSAPHSSSTSHLHHPNVAYDCLWNYSOYPSANPGNNLKDPLLSQFPGGQIPL 121
 Qy 127 NGILGSRQSPSSSHNTNLRAGSQKFWANGTHSPMGLNFDQSQELYDSDPPDNFE----- 180
 Db 122 NGILGSRQSPSSSHNTNLRAGSQKFWANGTHSPMGLNFDQSQELYDSDPPDNFEWMPNGP 181
 Qy 181 -----EVCSGIHPDAAAEKEMTSVAENGTGLVCSLELEE 216
 Db 182 PSFTSPQTSPLGSSIQTFAPSDVSDIHPDENAEEKELTSVAENGTGLVLSLELEE 241
 Qy 217 QPELKMCGYNGSVSPVESLHQEVSVLPDPPTVSCLDLDPDPSHLPDQLEDTPILSLESLEFPN 276
 Db 242 QPELKMCGYNGSVSPVESLHQEVSVLPDPPTVSCLDLDPDPSHLPDQLEDTPILSLESLEFPD 301
 Qy 277 SL-APEPVSGLYGIDDTFELMGAEDKPLXDSPVISALDCPSLNNATAFSLIADDSQTST 335
 Db 302 SLAAAEFVSGLYGIDDAELMGAEDKPLXDSPVISALDCPSLNNATAFSLIADDSQTSA 361
 Qy 336 SIFASPTSPVSGVLDQNSFDLNGSDAQEEMETQSSDFPPLSLTOPADQDQSTTQLH 395
 Db 362 SIFVSPSPVSGVLDQNSFGLNSCSDSEQEIEIETQSSNFQRPPLTEPAPDQPPSTQLH 421
 Qy 396 PATSPAVSPPTSPAVSLVSPAASPEISPEVCVPAASVTVSPAVSVSPASSAVLPAVSL 455
 Db 422 ----PAVSPTASPAASL-----TASAEISPAVSPVASSPVPEFVAVSPASSPALPAISL 473
 Qy 456 EVPLTASVTSKASPVTSAPAAFTASPAKDVSSFLETTADVEEITCEGLTASGSDVM 515
 Db 474 EASMTTPTVTSPOGSPSPAPAAFTQVSPAKDVSSAPKARADRETTGGAVAVSGSDVL 533
 Qy 516 RRRIATPEEVLPLQHGWRREVRIKKGSHRWQGETWYVPCGKMKQPEVIKYLRLV 575
 Db 534 KRRIATPEEVLPLQHGWRREVRIKKGSHRWQGETWYVPCGKMKQPEVIKYLRLV 593
 Qy 576 HSVREHFSFPRMPVGDFFERTDTEGLQVQLSAEIPSRIOAITKGRGRPRNTEKAK 635
 Db 594 HSVREHFSFPRMPVGDFFERTDTEGLQVQLSAEIPSRIOAITKGRGRPRNTEKAK 653
 Qy 636 TKEVPKVRGRGRPPKVKITELLNKTDNRPLKLEAQTNEEDKAKIAKSKKKMR 691
 Db 654 NKEVPKVRGRGRPPKVKITELLNKTDNRPLKLEAQTNEEDKAKIAKSKKKMR 709

RESULT 8
 BA2B HUMAN
 ID BA2B HUMAN STANDARD; PRT: 1972 AA.
 AC Q9UIF8; Q96EA1; Q96SQ8; Q9P252; Q9Y4N8;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 05-JUL-2004 (Rel. 44, Last annotation update)
 DE Bromodomain adjacent to zinc finger domain 2B (hwalp4).
 GN Name=BAZ2B; Synonyms=KIAA1476;
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RC TISSUE=Testis;
 RX MEDLINE=20130112; PubMed=10662543; DOI=10.1006/geno.1999.6071;
 RA Jones M.H., Hamana N., Nezu J., Shimane M.;
 RT "A novel family of bromodomain genes.";
 RL Genomics 53:40-45(2000).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM 3).
 RA Ohara O., Nagase T., Kikuno R.;

Db 369 SSQLPFIFQSSQAKEESVTKHTSVI--QSTGLVSNVKPLSLVNOAKKETRY----- 417
Qy 236 HQEVSVLVPDPTV-----SCLDDPSHLPDQLEDTPILSEDSLEPNSLAPEPVS----- 285
Db 418 ----KLIVPSDVLKAGKNVTSSESSLSLRSK---REQYKQTFPPQGGKKQEMGSLK 470
Qy 286 ----GLYGIDDTLMGABDKLPLXD-----SPVISAL--DCP-SLNN 321
Db 471 KVIAALSNTKATSSSPAHPKPLPDNNHPNPFITNALLGHQNGVQISVQIEAPLALTTK 530
Qy 322 TAFSLADDSQTSITSIFAS-----PTSPFVLGES----- 350
Db 531 TKMQSKINENVSSTPSPSPVNLSTSGRRAPGSGQTPALPSASPILHSSGKEKRVSNATP 590
Qy 351 -----VLQDNSFDLNGDAEQEEMETOSSDPPPLSLTQAPDQSTI 392
Db 591 LKAHHHPHAAAAALVQFGTSDVPSKDSNEDDEDEDEDEDEDEDEDEDEDESDS 650
Qy 393 QLHPATSPAVGPTTSPAVSLVSPAASPEISPEVCPAASTVVSFAVSVVSPASSAVLPA 452
Db 651 Q-----SESDSNSQDSEGEDDEKQDESDSDTEGEKPAV-NLTQTSSSAKSP- 699
Qy 453 VSLEVLPTASTVSPKASPVTSAAFPASPAKNDVSSFLETTADVEEITGEGLTAS-GS 511
Db 700 ----PSSLTAHSAPHLHIGSPPGSAPAAALCSESQPPAFLGTSSST-----LTSTPHS 747
Qy 512 GDVRRRIATPEEVRPLQHGWRREVRILKGSHRWQGETWYVYGGCGEMKQFPEVIKLS 571
Db 748 GTSKRRRVADDOELRIPLDYQWQRETRVRNFGGLRFGEVAYAFPGCKLRQCPDMVK--- 804
Qy 572 RNLVHSVRREHFSFSPRMPVDFEERTPEGLQWVQLSABEIPSRIOAITGKGRPRNT 631
Db 805 -----GMQCLLKEDVTPRIAMDGRGRPNP 833
Qy 632 EKAKTEKVPKVGGRGPRPKVKITELNKNTRPLKLEAQE-----TLNEED 679
Db 834 DRPRARESRMKRRGRPNVGSABFLDNTDAKLRLKLAQEIARQAQAIKLLRKLQKE 893
Qy 680 KAKTAKSKKKR-----QKVQRECLTTIQOARNKKEQETSKLKHKEAKKSKYAEK 733
Db 894 QARVAKEAKQQAIAAEBEKQKQEQMKIIOQEKIRKIQIR--MEKELRAQQILBEKE 951
Qy 734 GKTKQELKERVKEKKEKVRN-----KEKEEVTKAKPACADKTKLATQRLERQK 785
Db 952 LRRQAVLLKHQERERRQHVMLMKAMEARKKAEKELKQEKDEKELNKEKLEQRL 1011
Qy 786 QOMILEMKKPTEDMCLTDHQLPDPFSRVPGITLPSGAFSCLTIVPELHSGKVLGDP 845
Db 1012 ELEMAKELKPKEDMCLADQKLPFWPRIPGLVLSGTTTFDCLMVVQPLRNFVKVLGPDV 1071
Qy 846 AKDVPSLGVLOGLLCOGDSLGEVODLLVRLKALHDPGPSPYCSQSLKILGEKVSIEPL 905
Db 1072 NIDVFNLSVLOGLLNIIGDSGEVODLLVRLLSAAVCPDGLITGYKAKTALGEHLNNGV 1131
Qy 906 TRDNVSEILRFLMAYGVXPALCRLRTPQFQAQPPQKAAVLAFFPVHELNGSTLIINEI 965
Db 1132 NRDNVSEVLOLFMEAHCCQTELTSCLKYKAQAHTPAQKASILAFLVNELACSKSVSEI 1191
Qy 966 DKTLESMSYRNKRWIVGRLRLRLTKTLAKRTGRSE-----VEMGRPECLG-----RR 1014
Db 1192 DKNTEYMSNLRDKWVSGKRLKRI IHAKTKGRDASGGIDLGEHQPLGTPTPGRRR 1251
Qy 1015 RSSRIM-----EETSGNEEEEEEESIAAVPGRGR-----DGEVDATASSIPELERO 1062
Db 1252 RKGSDSDYDDDDDDSDQADEDEDEDEDDKDKGKKTDCICEDEDEGDQTA-SVEELEKQ 1310
Qy 1063 IEKLSKROLFPFRKLLHSSOMLRAVSLGQDYYRRYRWLPYLAGIFVEGTGCVLPVEEV 1121
Db 1311 IEKLSKQSQYRRKLPDASHLSRSMWFCQDYYRRYRWLPYLAGIFVEGTGCVLPVEEV 1369

RESULT 11
Q7Q3S9

ID Q7Q3S9 PRELIMINARY; PRT; 2060 AA.
AC Q7Q3S9;
DT 01-MAR-2004 (TrEMBLrel. 26, Created)
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE AGCP10911.
GN Name=agCS05231; ORFNames=ENSSANGG00000009298;
OS Anopheles gambiae str. PEST.
OC Eukaryota; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Anopheles.
OK NCBI_taxid=180454;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PEST;
RA Anopheles Genome Sequencing Consortium;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AAA01008964; EAA12387.1; -.
DR HSSP; Q9UIG0; 1562.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR001487; Bromodomain.
DR InterPro; IPR004022; DDT.
DR InterPro; IPR001739; Methyl-CpG_bind.
DR InterPro; IPR001965; Znf_PHD.
DR Pfam; PF00439; Bromodomain; 1.
DR Pfam; PF02791; DDT; 1.
DR Pfam; PF01429; MBD; 1.
DR Pfam; PF00628; PHD; 2.
DR PRINTS; PR00503; BROMODOMAIN.
DR PROSITE; PS00014; BROMODOMAIN_2; 1.
DR PROSITE; PS00827; DDT; 1.
DR PROSITE; PS01359; ZF_PHD_1; UNKNOWN_2.
DR PROSITE; PS00016; ZF_PHD_2; 2.
SQ SEQUENCE 2060 AA; 228520 MW; CBEF2DDFF7C95D40 CRC64;

Query Match 11.68; Score 1144; DB 2; Length 2060;
Best Local Similarity 21.14; Pred. No. 9.le-39;
Matches 459; Conservative 289; Mismatches 675; Indels 754; Gaps 63;
Qy 268 SEDSLEPNSLAPEPVSGLY---GIDDTLMGABDKLPLXDSPVISALDCPSLNATA 323
Db 57 SNDRIEVI-KLPPTITSGVYNACKGAGKDSMM-----DHNNWSG 96
Qy 324 FSLIA-----DDSTSTSI FASPTSPVGLSGVLODNGSFDLNGSDAQEEMETOSSD 376
Db 97 LNLKAGSGGGGGSSAAATIVSLSLTAGM-----NSSGNNQSSGMWQ-----DD 143
Qy 377 FPPSLTQPADQSTIQLHPATSPAVSPTTSPAVSLVSPAASPEISPEVCPAASTVV-- 434
Db 144 APLNLSM-KPSKSNEGMRSDSTHSAS-SSSPAWS-----GASANSLSQSLTITAA 192
Qy 435 -----SPAVF-----SWSPASSAV--LPAVSLEV---PLTASVTSPKA 468
Db 193 LGGNAGNANDSSRSQFEGEPRLNGRGVSKPKNTVASLLAQAQVSGSKPLTAQQLITQD 252
Qy 469 SPVTSAAAFTTSPANKDVSSFLETTADVEEITGEGLTASGGDVNRRRIATPEEVRLP 528
Db 253 AEIEKLQAQMLEASRNQSMNSTNTNTDTSISSESGMSESEGEHINKL-----ELRVP 307
Qy 529 LQHWRRREVIK---KSHRWQGETWYVYGGC-KRMKQFPEVIKYLNRNLVHSVRREHFS 584
Db 308 LEKGRWRETVIRGTRNGH-IKGDVYVYPPQSVNKMKGMMQIQILYDQFQPKDLNRNFS 366
Qy 585 FSPRMPVDFEERTPEGL--QWVQLSABEIPSRIOAITGKGRPRNTEKAKTEKVPKV 642
Db 367 FSAXAVGTFLQAPLPAFYATDGEFIKMTDVEVARLE----- 403
Qy 643 KRGRGRPPKVKITELNKNTRPLKLEAQETLNEEDKAKIAGKMKRQKVGECGLTT 702
Db 404 -----DLKMFTRHAGLVEQRIEIAKQQQALR----- 430

DR	PROSITE; PS50016; ZF PHD 2; 1.	
FT	NON TER 796	
SQ	SEQUENCE 796 AA; 88462 MW; 2EDB8BCAF9720C1E CRC64;	
	Query Match	
	Best Local Similarity 11.3%; Score 1108.5; DB 2; Length 796;	
	Matches 298; Conservative 124; Mismatches 283; Indels 187; Gaps 33;	
QY	972 MSSRYKNWIVGRLRRRLKTVLAKTGRSE-----VEMGRPECLG-----RRSSRIM 1020	
DB	1 MSNLRDKWVVEGKRLKRIIHTAKTKGRDTSGGIDLGEEQHPLGTPPTGPKRRRKGGDS 60	
QY	1021 E-----ETSGMEEEEESIAAIVGRGR--RDGEVDATASSIPELRQTEKLSKR 1069	
DB	61 DYDDDDDDSDGDEDEDEDEKEDKXKTKTDICEDDEGDAQASVELEKQLEKLSKQ 120	
QY	1070 QLFPRKLLHSSQMLRAVSLGODRYRRRYVLPYLAGIFVETGTEGNLVPVIVIKETDSL 1129	
DB	121 QSVRRKLFDAHSLSRSLRVFMFGDQRYRRRYWILPQCGGIFVEGMSGESGLEE-IAKEREL 179	
QY	1130 KVAHASLNPALFSKMLASNTTASSPARSR-PLKTPG----FMQ-PRHFKSPVRG 1184	
DB	180 KKAESVQIKEMF-----ETSGDLSNCSNTDHCQEKDLKEDKNTNLFQKPGFSFKLSKL 235	
QY	1185 ODSEQOQAOLQPAOLHVPAPQPOLQLOSHKGFLQEGLSPLSGOS--QHDLS--QSA 1241	
DB	236 LE-----VAKMPESEVWP-----KPNAGANGCTLSYQNSGKHSLGSVQS- 276	
QY	1242 FLSWLSOTQSHSSLLSSVL--TPDSSPGKL-DPAPSQPPEPEDEAESPDLPQAFWN 1298	
DB	277 ----TATQSNVEKADSNLNTGSSGPGKFVSLPND-----QLLKTUTEKNRQWFS 324	
QY	1299 ISAQWPCN-----RAPTPLAVSDEQTPSPQQLASSKPMRPSAANP--CSP 1344	
DB	325 LLRPTCDPTSLTHADMSTASLVTPQSPQSPKSPSPTEPAPLGSS--AQNPVGLNPFALSP 382	
QY	1345 VQF--SSTPLAGLAKRRAGDP-----GEMPOSPTGLGQPKRRG-----PPS 1385	
DB	383 LQVKGVSVMGL---QFCGWITGVVTSNIPITSSVPSLGSGLGISEGNGSNFSLTNVASS 439	
QY	1386 KF---FKOMEQRYLTQLTA-----QVPPPEMCSGWWIIPDEMLDAMLKALHP 1430	
DB	440 KSESVPQNEKATSAQPAAVEVAKVPVDFPSPKPIPEEQFGWRIIDPEDLKALLKVLHL 499	
QY	1431 RGIREKALHKLHKRDFLOEVLCLRPSADTPIEPQLPAFO---EGMSWSPKKTETD 1487	
DB	500 RGIREKALQKIQKHLDYITQACLKNKDVAIIELNENEENQVTRDIVENWSVEEQAMBMD 559	
QY	1488 LAVLQWVEELQORVIMSDLIQRTGTCPSDPSTREDLAYCEHLSDS---QSDITWRGPR 1543	
DB	560 LSVLQVEDLERRVNASLQVKGWMCPEPASEREDLVYFEKSTKLCKEHGDGFTGEDE 619	
QY	1544 EGL-APQKTTNPLDLAVMRLAALQNVKRYLRLEPLWTHVLEKALLSTPNPAGPT 1602	
DB	620 SSAHALERKSDNPLDIAVTRLADLERNIER-----650	
QY	1603 TTETSYETPIRIRWQTLQRCSAAHVCLGLHLERSIAWEKSNVKTCLVCRKGDND 1662	
DB	651 ---FEEDIAPGLRVRRALSEARSAQVALCIQQLOKSIAWEKIMKYQICIRKGDNEE 707	
QY	1663 FLLLCDGDCRGCHYICHRPKMEAVPEGDWFTVCLAAQVGEFTQKPGFPRGQKRKSGY 1722	
DB	708 LLLLCDGCDCKGCHYICHRPKITTPIDGDNWFCFACIA-----KASGQ 748	
QY	1723 SLNFSE--GDGRRRLVLLKGRESAPAGPRYSEERLSPSKRRRLSMRNNHSDL 1772	
DB	749 TLKIKKLVHKGKKTNESKKGKVKTLTGTDTEDEDSASTSS----SLKRGNKDL 796	
RESULT 13		
Q9V602		
ID	Q9V602	
AC	Q9V602;	

DT	01-MAY-2000 (TrEMBLrel. 13, Created)	
DT	01-OCT-2002 (TrEMBLrel. 22, Last sequence update)	
DT	01-MAR-2004 (TrEMBLrel. 26, Last annotation update)	
DE	CGI0897-PA.	
GN	Name=tou; ORFNames=CGI0897;	
OS	Drosophila melanogaster (Fruit fly).	
OC	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;	
OC	Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;	
OC	Ephydroidea; Drosophilidae; Drosophila.	
OX	NCBI_taxID=7227;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RP	MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;	
RA	Adams M.D., Celnikier S.E., Holt R.A., Evans C.A., Gocayne J.D.,	
RA	Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,	
RA	George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,	
RA	Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,	
RA	Brandon R.C., Rogers Y.H., Blazee R.G., Champe M., Pfeiffer B.D.,	
RA	Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,	
RA	Abril J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,	
RA	Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,	
RA	Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,	
RA	Borkova D., Botchan M.R., Bouch J., Brokstein P., Brottier P.,	
RA	Burtis J.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,	
RA	Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,	
RA	de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,	
RA	Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,	
RA	Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,	
RA	Fowler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,	
RA	Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,	
RA	Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,	
RA	Hoslin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,	
RA	Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,	
RA	Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,	
RA	Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,	
RA	Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,	
RA	Metkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,	
RA	Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,	
RA	Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,	
RA	Palazzo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,	
RA	Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,	
RA	Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,	
RA	Spieler E., Spradling A.C., Stapleton M., Strong R., Sun E.,	
RA	Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,	
RA	Wang Z.Y., Wassarman D.A., Weinstock G.M., Weissbach J., Ye J.,	
RA	Williams S.M., Woodage, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,	
RA	Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,	
RA	Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,	
RA	Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;	
RT	"The genome sequence of Drosophila melanogaster.";	
RL	Science 287;2185-2195(2000).	
RN	[2]	
RP	SEQUENCE FROM N.A.	
RP	MEDLINE=2426065; PubMed=12537568;	
RA	Celniker S.E., Wheeler D.A., Kronmiller B., Carlson J.W., Halpern A.,	
RA	Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,	
RA	George R.A., Hoskins R.A., Lavery T., Muzny D.M., Nelson C.R.,	
RA	Pacleb J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,	
RA	Svirskas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,	
RA	Weinstock G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.;	
RT	"Finishing a whole-genome shotgun: Release 3 of the Drosophila	
RT	melanogaster euchromatic genome sequence.";	
RL	Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).	
RN	[3]	
RP	SEQUENCE FROM N.A.	
RP	MEDLINE=2426070; PubMed=12537573;	
RA	Kaninker J.S., Bergman C.M., Kronmiller B., Carlson J., Svirskas R.,	
RA	Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,	
RA	Ashburner M., Celniker S.E.;	
RT	"The transposable elements of the Drosophila melanogaster euchromatin:	
RT	a genomics perspective.";	
RL	Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).	
RN	[4]	

QY 486 -----KDVSSSEL-----ETTADVEITGEGLTASGSDVMRRRIA----- 520
Db 264 EODVKNKQOEELTKLMAHQJLHOKQKEKKQADAAALAAKVLAA-----HRAALES 315
QY 521 -TPEE-----VRLPLQHGWRREVRIKK-GSHRWQGETWYTCGKRMKQFPPEVIKYL 570
Db 316 DSPSEGGKKTNEAWLRLPLQHGWRRTCVRSIASAGVKGDSYFAPCGKKLSTYSSEVRYL 375
QY 571 SNLVHSVVRREHFSRPMVGDFFPEERDTPGLOVQVLSAEIIPSRQAITKGRGRPN 630
Db 376 TKNSHYITRDNFENTKLVIAGEFT----- 400
QY 631 TEKANTKEVPKVRGRGRPPVKITELNKTNDNRPLKLEAQETLNEEDKAKIAKSKKM 690
Db 401 -----VP-----KQTEADETOQEREFAMF--TEDDI 424
QY 691 ROKVORGECLTTIOGOARNKRKQETSLKHKEAKKSAEKEKTKQEKLEKVKREKK 750
Db 425 NKELTNLNVLKVP-----KIOASTSNGVHEDDIKMSKIEEPDEPLDPSSELNDEFER-- 477
QY 751 EKVNMKEKEVNTAKPACKADKTLATQRRLERQKQMILEBMMKPTEDMCLTDHQPPLD 810
Db 478 ----LVHSQIMSGVDECK-----IREREADLLVN-----INDVRHLPD 513
QY 811 FSRVPGTLTPSGAFSDCLTIVEFLHSFGVLGFDPAKDVPSLGLVQBLGCGDLSGEVQ 870
Db 514 PSRIGNQCLSSQGFADALMVHFVQNGHVLGD--LEIAPKLESACAGLDGDANHAQTL 572
QY 871 DLLVLLKAAALHDGPPSYCQSLKILGEKVSPIPLTRDNVSEILACFLMAYGV-XPALCD 929
Db 573 QLTRQLLLALALEFPQNGNE----KFPQGGGEMGLDRENFSEVMFLFLIDKGRGEELSQ 628
QY 930 RLRTOPFOAQQPQKAAVLAPFVPHELNGSTLIINEIDKLTLESMSYRKKNKIVIGELRLRL 989
Db 629 PLLTNCNFLSISPEQASILAFCLDELVCVRNVVTEIDKNLDEISRLKGEKMWREKARAL 688
QY 990 KTVLAKTRGRSEVENGRPEELGRRSRIMEETSGMEBEESIAAVPGRGRDRGEV 1049
Db 689 RSARSKKKNDKVVV-----VKEEQNHESDSEPP---TRPDTPK 724
QY 1050 DATAS-----SIPELERQIEKLSKRLQF 1072
Db 725 KATVAPPTVSVSVSAACQOQRKFTPGLGCEVLTEQESMSLQOQMSLDGLDHOEAQN 784
QY 1073 FRKLLHSSQMLRAVSLGQDRYRRYVWLPYLAGIFVREGTEGNLVPVEVIKKTDLSKVA 1132
Db 785 INQKIHDTGLKIRSPFGTDRHRYWMLAHTDKVIE-----SLA 825
QY 1133 AHASINPALFSMKMELAGSNTTASSPARARSRLTKPGFWQPHFKSPVRQDSEQPOA 1192
Db 826 TTSVNNPA-----CNANEVAS-----KDP----- 844
QY 1193 QLQPEAQLHVPAQPOQLQLQSHKGFLEQEGSPLSLQSOHDLSQSAFLSWLSQTOH 1252
Db 845 ---PTEORVPG----- 853
QY 1253 SLLSSSVLTDPSSPGKLDPAPOPPPEPEDEABESSPLQAFWFNISAQWPCNAAPTTP 1312
Db 854 -----ACETIDLDV 862
QY 1313 LAVSEDOPTPSPQQLASKPMNRPSAANPCSPVQFSSTPLAGLAPKRRAGDPGEMPOSPT 1372
Db 863 IACVED-----LVDDVVLRL----- 877
QY 1373 GLGQPKRGRPPSKFFKQMEQRYLTQTAQPVPPPMCSGWWIPDPEMLDAMLKALHPRG 1432
Db 878 -----AKADKTKRKRY-----RRINEMKRGWTCMQNRDCVESLRSCLMSRG 919
QY 1433 IREKALHKLHNRDQFLQEVCLRPSADIFEPRLQPAFQEGIMSWSPKEKTYETDLAVLQ 1492
Db 920 IREKALHLLTK-----PWFLNELKFGTIT----- 944
QY 1493 WVEELEQRVIMSDLQI---RGWTCPSDPTREDLAYCEHLSDSQSDITWRGPGREGAPQ 1549

Db 945 -IEPVGEK---SDLELVRROQMT---RLNTAIDKLQC-HLKMS--DVSKPLPSITPFTQ 994
QY 1550 RKTNTPLDLAVWRLA-----ALEQNVRKRYLREPLWPTHEVLEKALLSTPNGAPECT 1602
Db 995 KPIVVPPTMALAQIVKDDMAKVDEEDVGQELDE-----TIROKLI-----E 1038
QY 1603 TTEISYEITPRIRIWRQTLQ-RCRSAHVCLGLHLSIAWKSXVNVKTVCLVCRKGDND 1661
Db 1039 TADMDDQTCQLFEDWKSYSTEAQTTSQMLVALQTLQEGIMWRSREALCQICKSMDGD 1098
QY 1662 EFLLLDCGDRGCHLYCHRPKMEAVPEGDWCTWC----- 1696
Db 1099 E-MLVDCDESGCHMECFRPMTKVPEGDWFCQREBSGRPMCMFCRSRETGNLHQCOR 1157
QY 1697 LAQOVGEFTQ-----KGFPRK----- 1715
Db 1158 CAYHVHQESODGPKKEALNPETFCGHCQEMKQKRFVZKLLILRSESEBELEDNDHAENG 1217
QY 1716 QKRSGVSLNPSSEGDRRRRVLLKGRBSPAAGPRYSEERLSPSKRRRLSM-----RNH 1768
Db 1218 ENTNGH-MNGMNG-----AIAIGVHNOQNGVKGNLKRKLEVPISIGGLPKNM 1263
QY 1769 HSDLTFCCEIILMEMESHDAAMPFYEPVNPRLVSGYRRIIKNPMDFSTWRERLLRGYTS 1828
Db 1264 NKEL--COLMDELVQANALPFLPEVNPGLVPGYKMLISKPMDLKTIROKNEKLIYETP 1321
QY 1829 EEFADALLVDCNCTFNEDDSEVGKAGHIMRRRPFESWEE 1869
Db 1322 EDFADIELMPANCROFNIDHSEIGRAGISLHKFFQKRWKQ 1362

RESULT 16

Q8BRP6 PRELIMINARY; PRT; 181 AA.
ID Q8BRP6
AC Q8BRP6
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2004 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Mus musculus 10 days neonate cortex cDNA, RIKEN full-length enriched library, clone:A830034W21 product: bromodomain adjacent to zinc finger domain, 2A, full insert sequence. (Fragment).
DE domain, 2A, full insert sequence. (Fragment).
GN Name=Baz2a;
OS Mus musculus (Mouse).
OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Cortex;
RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning.";
RL Meth. Enzymol. 303:19-44 (1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Cortex;
RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RA RIKEN FANTOM Consortium;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690 (2001).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Cortex;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs.";
RL Nature 420:563-573 (2002).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Cortex;
RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxId=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM 1), AND MASS SPECTROMETRY.
 RC TISSUE=Cervical Carcinoma;
 RX MEDLINE=20341093; PubMed=10880450; DOI=10.1093/emboj/19.13.3377;
 RA Poot R.A., Delaire G., Huelsmann B.B., Grimaldi M.A., Corona D.F.V.,
 RA Becker P.B., Bickmore W.A., Varga-Weisz P.D.;
 RT "HUCHRAC, a human ISWI chromatin remodelling complex contains hACFI
 and two novel histone-fold proteins.";
 RL EMBO J. 19:3377-3387(2000).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RX MEDLINE=20122567; PubMed=10655480; DOI=10.1073/pnas.97.3.1038;
 RA Bochar D.A., Savard J., Wang W., Lafleur D.W., Moore P., Cote J.,
 RA Shkhattar R.;
 RT "A family of chromatin remodeling factors related to Williams syndrome
 transcription factor.";
 RL Proc. Natl. Acad. Sci. U.S.A. 97:1038-1043(2000).
 RN [3]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RC TISSUE=Testis;
 RX MEDLINE=20130112; PubMed=10662543; DOI=10.1006/geno.1999.6071;
 RA Jones M.H., Hamana N., Nezu J.-I., Shimane M.;
 RT "A novel family of bromodomain genes.";
 RL Genomics 63:40-45(2000).
 RN [4]
 RP SEQUENCE OF 1-541 FROM N.A. (ISOFORM 1).
 RC TISSUE=Blood;
 RA Ye M., Zhang Q.H., Zhou J., Shen Y., Wu X.Y., Guan Z.Q., Wang L.,
 RA Fan H.Y., Mao Y.F., Dai M., Huang Q.H., Chen S.J., Chen Z.;
 RT "Human partial CDS from cd34+ stem cells.";
 RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE OF 746-1556 FROM N.A. (ISOFORM 1).
 RC TISSUE=Uterus;
 RA Wambutt R., Heubner D., Mewes H.-W., Gassenhuber J., Wiemann S.;
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: May play a role in transcriptional regulation. May be
 involved in the formation or maintenance of heterochromatin
 playing a critical role in developmental control.
 CC -!- SUBUNIT: Together with CHRAC1, POE3 and ISWI/SNF2H proteins, it
 forms the ISWI chromatin remodelling complex, CHRAC.
 CC -!- SUBCELLULAR LOCATION: Nuclear. May target the CHRAC complex to
 heterochromatin.
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event-Alternative splicing; Named isoforms=2;
 CC Name=1;
 CC IsoId=Q9NRL2-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=Q9NRL2-2; Sequence=VSP_000551;
 CC -!- TISSUE SPECIFICITY: Highly expressed in testis and at low or
 undetectable levels in other tissues analyzed.
 CC -!- MISCELLANEOUS: Stimulated by double-stranded DNA and nucleosomal
 DNA.
 CC -!- SIMILARITY: Belongs to the WAL family.
 CC -!- SIMILARITY: Contains 1 bromodomain.
 CC -!- SIMILARITY: Contains 1 DDT domain.
 CC -!- SIMILARITY: Contains 1 PHD-type zinc finger.
 CC -!- SIMILARITY: Contains 1 WAC domain.
 CC -!- CAUTION: Ref.4 sequence differs from that shown due to multiple
 frame shifts starting at position 382.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF213467; AAF70601.1; -;
 DR EMBL; AF221130; AAF32366.1; -;

DR EMBL; AB032252; BAA89209.1; -;
 DR EMBL; AF161435; AAF28995.1; ALT_SEQ.
 DR EMBL; AL050089; CAB43261.1; -;
 DR PIR; T08738; T08738.
 DR HSSP; Q9UIG0; 1P62.
 DR Genew; HGNC:960; BAZ1A.
 DR MIM; 605680; -;
 DR InterPro; IPR001487; Bromodomain.
 DR InterPro; IPR004022; DDT.
 DR InterPro; IPR001965; Znf_PHD.
 DR InterPro; IPR001841; Znf_ring.
 DR Pfam; PF00439; Bromodomain_1.
 DR Pfam; PF02791; DDT; 1.
 DR Pfam; PF00628; PHD; 1.
 DR PRINTS; PR00503; BROMODOMAIN.
 DR SMART; SM00571; DDT; 1.
 DR SMART; SM00249; PHD; 1.
 DR SMART; SM00184; RING; 1.
 DR PROSITE; PS00633; BROMODOMAIN_1; 1.
 DR PROSITE; PS50014; BROMODOMAIN_2; 1.
 DR PROSITE; PS50827; DDT; 1.
 DR PROSITE; PS01359; ZF_PHD_1; 1.
 DR PROSITE; PS50016; ZF_PHD_2; 1.
 KW Alternative splicing; Bromodomain; Coiled coil; Nuclear protein;
 KW Transcription regulation; Zinc-finger.
 FT DOMAIN 22 128 WAC.
 FT DOMAIN 422 487 DDT.
 FT ZN_FING 1148 1198 PHD-type.
 FT DOMAIN 1446 1516 Bromodomain.
 FT DOMAIN 306 397 Coiled coil (Potential).
 FT DOMAIN 634 709 Coiled coil (Potential).
 FT VARSPLIC 504 535 Missing (in isoform 2).
 FT /FTid=VSP_000551.
 FT CONFLICT 135 135 R -> T (in Ref. 3).
 FT CONFLICT 236 236 D -> E (in Ref. 3).
 FT CONFLICT 551 551 D -> E (in Ref. 2 and 3).
 FT CONFLICT 730 730 V -> F (in Ref. 3).
 FT CONFLICT 769 769 P -> L (in Ref. 3).
 FT CONFLICT 1201 1201 S -> C (in Ref. 3).
 FT CONFLICT 1206 1206 S -> F (in Ref. 3).
 FT CONFLICT 1366 1366 K -> N (in Ref. 2 and 3).
 FT CONFLICT 1409 1409 K -> R (in Ref. 2 and 3).
 SQ SEQUENCE 1556 AA; 178674 MW; 43F2667C3CFD458 CRC64;

Query Match 6.7%; Score 660.5; DB 1; Length 1556;
 Best Local Similarity 19.7%; Pred. No. 5.4e-19;
 Matches 307; Conservative 198; Mismatches 411; Indels 641; Gaps 45;

Qy 644 RGRGRPPKVTITLLNKTNRPLKLEBAQETLNEDRAKIAKSKKMRQKVRQGECLPTTI 703
 Db 274 RRRGRPPK-RIH--ISQEDN-----VANKQTLSYRSKATKERDKLLKQEMKSL 320
 Qy 704 QGQARNKRKQETSLKHKAKKSAEKEGKTQKQKLKEKVKREKKEKVKMKKEEVTK 763
 Db 321 AFEXAKLKREKADAL---EAKKKEKEDK---KREELKKTIVEE---RLKKKEE--- 366
 Qy 764 AKPACKADKTLATORLEERQKQOMILEEMKPTEDMCLTDHQLPDPFSRVPGLTLPGA 823
 Db 367 -KERLKYEREKREKREKREKRYEYLKQWSKPREDECDLDELPEPTPK-TRLPPEI 424
 Qy 824 FSDCLTIVFHLSPGKVLGPDPAKDV---SLGVLOEGLLCQGSLSGVODLLVRLKKA- 879
 Db 425 FGDALMVLFLNARFEL--FDLQDFPDGVTLEVEEALV-GNDSGGLPCELLFFLTAI 481
 Qy 880 -----ALHDPGFP-----SYCSLKIL-----GEKYSV 902
 Db 482 FQATAESEEEVAKESQLTDADTKDTEALDEADPTKSALSASVASLAARAWPOLHQCSLKS 541
 Qy 903 IPLTRDNVSEILRCFLMAYGV-----XPALCDRLR 932
 Db 542 LDLDSCITSLDLRLHLASGADVTSANAKRYKRGGFDTDDACMELRLSNPSLVKKLS 601

Qy	933	TQPFQAQPOQKA	AVLAFPHVH	NGSTL-----	LINEIDK	TLESMS--	YR	976
Db	602	STSVYDITPE	GKMKIL-----	HALCGKLL	TLTVST	DFIEDY	VDIL	ROAKQEFRELKABQHR
Qy	977	KNKWI	VEGRLLRLK	TVLAK-----	RTGRSE	VEMG	GRPE----	ECLGRR
Db	658	KERE	AAARIRK	KEKKEK	QEQKQK	QKQK	KEDEQR	NSTADISIGEE
Qy	1015	RSR	MEETS	MEEBEE	EEESIA	AVPGR	RGR--	DGEVDAT-----
Db	718	KDTE	QEXELD	QDMVTE	DEDDPG	SHKGRG	KRGQ	GFKEFTQEQIN
Qy	1066	LSK	QLFPR	KKLH	SHSOM	LRAVS-----	LGOD	RVRRRYVVL
Db	778	LKQ	EHQK	EKELLE	KIQSA	IACTNI	PELGR	DRMYRWIFPS
Qy	1117	VP	EVIK	ETDSL	KVA	HAASIN	PALF	PMKMLAGSNT
Db	838	LP						
Qy	1177	HF	KSPVR	QDSE	QEQPA	QLQPEA	QLHVP	QPOQLQLQ
Db	843	SF	QNV	QSD-----	PQV-----			
Qy	1237	LS	OSAF	LSWL	SQTS	SHSS	LS	SSVLTDP
Db	856							
Qy	1297	FN	ISAQ	MP	CNAAPT	PPPLA	VSED	QTPPS
Db	877	DH	SVQ	LP-----				
Qy	1357	PK	RAAG	DG	PM	QSP	TGLG	QKRRG
Db	892							
Qy	1417	DP	EMLD	AML	KALH	PRG	IRE	KALH
Db	897	SCE	LD	QLI	EALN	SRGH	RESA	LKETT
Qy	1472	EG	MS--	WSP	KE	TYE--	TD	LAVQW
Db	957	RGR	SSN	AYD	PSQ	MAEQ	LE	RLRDF
Qy	1513	CP	SD-----	STRE	LAV	CEHL	SD	QBDIT
Db	1017	LL	SEN	KENG	IIT	VNE	D	EMELD
Qy	1556	LD	LAV	MLA	LA	LEQ	NK	RYLREPL-----
Db	1077	VH	LAM	AL	FQ	IEQ	IERR	FLKAP
Qy	1611	TP	IRI	WR	QTL	QRC	SAH	VC
Db	1116							
Qy	1671	DR	GHI	CH	RP	KMA	VP	EGD
Db	1170	DR	GHTY	C	VR	PK	LT	VP
Qy	1709	PG	FK	QK	QK	SG	YS	LN
Db	1230	DG	DE	EQ	S	EE	E	V
Qy	1743	SP	AAG	PR	YSE-----			
Db	1290	EP	GR	PS	R	S	Q	S
Qy	1753	ER	LS	P	S	K	R-----	
Db	1350	EL	L	S	P	R	K	R
Qy	1771							

Db 1410 KROSPBPSTVTLGRRSSGRQGVHEUSAFELQVLVLRHDDSSWPFLLKLVSKIQVDPYDI 1469
QY 1807 IKNPMPDFTNRRLLRGGYTSSEFAADALLVFDNCOTFNEDDSEVGKAGHIMRRFF 1863
Db 1470 IKPIALNIIREKVNKCEYKLASEFDDIITELMESNCFEYNPRNTSEAKATRLQAFF 1526

RESULT 19

BALB_MOUSE

ID BALB_MOUSE STANDARD; PRT; 1479 AA.

AC Q9Z277; Q9CU68;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Bromodomain adjacent to zinc finger domain protein 1B (Williams-Beuren
syndrome chromosome region 9 protein homolog) (WBRS9).
GN Name=Baz1b; Synonyms=WBSrcr9;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]

SEQUENCE FROM N.A.
RP MEDLINE=9907764; PubMed=9858827;
RX Medline=C57BI/6J; TISSUE=Thymus;
RC STRAIN=C57BL/6J; DOI=10.1038/nature01266;
RX MEDLINE=22354683; PubMed=12466851; DOI=10.1038/nature01266;
RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
RA Nikatido I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,
RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,
RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
RA Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,
RA Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,
RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,
RA Gaestland T., Gariboldi M., Giassi C., Godzik A., Gough J.,
RA Grimmeron S., Guetincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,
RA Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,
RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,
RA Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,
RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,
RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,
RA Sandelin A., Schneider C., Sample C.A., Setou M., Shimada K.,
RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,
RA Varaldo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,
RA Wilming L.G., Wyshaw-Boris A., Yanagisawa M., Yang I., Yang L.,
RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,
RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,
RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,
RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,
RA Miyaazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,
RA Yaeunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
RA Birney E., Hayashizaki Y.
RA Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs."
RL Nature 420:563-573(2002).

CC -I- FUNCTION: Forms a chromatin remodeling complex that mobilizes
CC nucleosomes and reconfigures irregular chromatin to a regular
CC nucleosomal array structure (By similarity).
CC CC -I- SUBUNIT: Interacts with ISWI (imitation SWI protein) to form the
CC WSTF-ISWI chromatin remodeling complex (WICH) (By similarity).
CC CC -I- SUBCELLULAR LOCATION: Nuclear (Potential). Accumulates in
CC pericentromeric heterochromatin during replication (By
CC similarity).
CC CC -I- TISSUE SPECIFICITY: Expressed in all tissues examined including
CC heart, brain, spleen, lung, skeletal muscle, kidney and testis.
CC CC -I- DEVELOPMENTAL STAGE: Expressed as early as day 7 and in equal
CC

CC amounts during gestation.
 CC -!- SIMILARITY: Belongs to the WAL family.
 CC -!- SIMILARITY: Contains 1 bromodomain.
 CC -!- SIMILARITY: Contains 1 DDT domain.
 CC -!- SIMILARITY: Contains 1 PHD-type zinc finger.
 CC -!- SIMILARITY: Contains 1 WAC domain.
 CC -----
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AF084480; AAD08676.1; -;
 CC EMBL; AK017894; BAB30992.1; -;
 CC PIR; T17401; T17401.
 CC HSSP; Q9UG0; 1F62.
 CC MGD; MGI:135499; Bazlb.
 CC GO; GO:0005721; C:centric heterochromatin; IDA.
 CC GO; GO:0000793; C:condensed chromosome; IDA.
 CC GO; GO:0005634; C:nucleus; IDA.
 CC GO; GO:0005515; F:protein binding; IPT.
 CC GO; GO:0006333; P:chromatin assembly/disassembly; IDA.
 CC GO; GO:0006338; P:chromatin remodeling; IDA.
 CC InterPro; IPR001487; Bromodomain.
 CC InterPro; IPR004022; DDT.
 CC InterPro; IPR001965; Znf PHD.
 CC Pfam; PF00439; Bromodomain; 1.
 CC Pfam; PF06628; PHD; 1.
 CC PRINTS; PR00503; Bromodomain.
 CC SMART; SM00297; BROMO; 1.
 CC SMART; SM00571; DDT; 1.
 CC SMART; SM00249; PHD; 1.
 CC PROSITE; PS00633; BROMODOMAIN_1; FALSE_NEG.
 CC PROSITE; PS50014; BROMODOMAIN_2; 1.
 CC PROSITE; PS50827; DDT; 1.
 CC PROSITE; PS01359; ZF PHD 1; 1.
 CC PROSITE; PS50016; ZF PHD 2; 1.
 CC Bromodomain; Coiled coil; Nuclear protein; Transcription regulation;
 CC Zinc-finger. 20 126 WAC.
 FT DOMAIN 605 669 DDT.
 FT ZN FING 1184 1234 PHD-type.
 FT DOMAIN 1352 1422 Bromodomain.
 FT DOMAIN 537 587 Coiled coil (Potential).
 FT DOMAIN 769 815 Coiled coil (Potential).
 FT DOMAIN 851 895 Coiled coil (Potential).
 FT DOMAIN 1257 1284 Coiled coil (Potential).
 SQ SEQUENCE 1479 AA; 170788 MW; DFB5816AE1C0634F CRC64;
 Query Match 5.5%; Score 538; DB 1; Length 1479;
 Best Local Similarity 18.8%; Pred. No. 5.7e-14;
 Matches 352; Conservative 249; Mismatches 624; Indels 646; Gaps 70;
 QY 141 HNTNLRAGSQKFWANGTHSPMGLNF-----DSQE----LYDSFPDNPFEVCSGHPDEA 191
 DB 38 YEALERYSERIWTCKSTGSSQLTHKEAWEEQEAELKKEFPNWEKLVLEWHHTA 97
 QY 192 AEKWT-----SVAENGTGLVCSLEL-BEQPELKMCGYNGSVPSLSHQEVSVLVD 245
 DB 98 SLEKLVDASLWEILMTKYAVGECDPEVGKMKLVKIV-----KHPLEKVDDEA---VEK 150
 QY 246 PTVSCLDPPSHLPDQLEDTPLSEDSLEFPNLSLAPVPVSGLYGIDD-----TEL 295
 DB 151 KSDGACDSFS---SDKNSSQMAQD-LQKQETVVKED-EGRESINDRARSPRKLPTSL 205
 QY 296 MGAEDK-----LPLXDSFVIALDCPSLNNATAFSLLDSDTSIFASPTSPVLGE 349
 DB 206 KGERKWAAPPKLP-----HKYDVKLQNEDKIISNVPADSLIRTE---RPPNKEIL-R 254
 QY 350 SVLQDNSFDLNNGSDAE---QEEMETQSDFFPSPSLTQAPDQSDSTIQLHPATSPAVSPTT 406

DB 255 YFIRHNALRAGTGNAPWVVEDELVKKYS-----LPSKFSDFLDPYKYMTLNFST 305
 QY 407 SPAVSLVSPAPASPEISPEVCVPAASTVVS PAVFSVVS PAVSGLVPAVSLEVPPLTASVTSP 466
 DB 306 KR-----RNTGSPDRKPSKPKRDS-----SSLSSPLNPKLVCHVHLEKSLNGP 349
 QY 467 -----KASPTVSPAAAFPTTASPANKDVSSFLETTADVEETGEGTLTASGSDVMRR 517
 DB 350 PLKVNKSNKSPBEHLEGVNKMIMSPNNKLHSP-----HIPKGPAAK----- 393
 QY 518 RIATPEEVRLPLQHCWRRREVIKYGSHRWQSTWYCGCGKRMKQFPEVIKVLSENVLVS 577
 DB 394 -----KPKKH-----SDKPLKAKGRGRGILNG 415
 QY 578 VRREHFSRPMVGDFFPERDTPGQLQWVQLSABEIPSRQAIT-----GKRGRPRNTEK 633
 DB 416 QKSTGNSKSPSKCV-----KTPK-TVMKQMTLLDMAKTQKMTTRPSRSGGVSRSGK 467
 QY 634 AKTKEVPKVGKGRGPRPKVITELNLTNDNPLKK-----LEAQTLENEEDKAKIAKS 686
 DB 468 PH-KHLP-----PAALHLIAVYKKNKDKEDKKSALS CVISKRTARELSNEDRARLPEE 518
 QY 687 KKKVRQKVGQCECLTTTQGOARNKPKQETKSLKHKE--AKKSXAEEKGKTKQSKLKEK 744
 DB 519 LRALVQK-----RYELLEHKRWASMSSEBQRKEYLKKKROELKER 558
 QY 745 VKRSKKEVKMKKEEVTAKPACKADKTATQRLERERQKQMLEBMMKPTEDMCLTD 804
 DB 559 LREKAKER---REREMLER-----LEKQKRFEDQE-----LG 587
 QY 805 HQPLPDFSRVFGLT-LPFGAFSDCLTIVFELHSGKVLGFDPAKDVPSLVQLQEGLLCQG 863
 DB 588 GRNLPAPFLVDTPGELPNTLFGDVALVVEFLSCYSGLLPDAQVPI TAVS-LMEALSADK 646
 QY 864 DSLGEVDLLVRLKAAHDEGFPFSYCSLKILEKVSIEIPLTRDNVSEIILRCFLMAYG- 922
 DB 647 GGFLYLARVILLQTLLODEIAEDYGE---LGMKLSIEIPLTTHSVSELVRLCLRRCDV 702
 QY 923 -----VXPALCDRLRTPFQAQPPQQAALVAFPVHLENGSTL 960
 DB 703 QEDSEGSETDNKDSPTFEDNEVDQEFLEKLETFEFTSEKRLITALCHRI-----L 758
 QY 961 IINEIDKTLSEMSYRNKNWIVEGRRLRK----TVLAKRTGRSEVEN-----GRPERCL 1011
 DB 759 MTYSVQDHMETRQQVSAELW--KERLAVLKEENDKGAEKQKREMEARNKENGKEENVL 816
 QY 1012 GRRSSRIMEETSGMEEBEESIAAUPGRR-----GRDGEVDATASSIPELEKQIKL 1066
 DB 817 GKVDKKEIVKIEQQVEEADDMISAVKSRRLLSMQAKRKEIOERETKV-RLEREAESE 875
 QY 1067 SKR-----QLFFRRKLLHSSQMLRAVSLGODYRRRYVWLP-VLAGIFVSGTEGNLYPE 1119
 DB 876 RMRKHKAAAEKAFQEGGAKAKLVLRRTPTIGTRNHNRYLWFSNEVPGLFIE----- 926
 QY 1120 EVIKKETDSLKVAHAHSLNPAFLFSMKMELAGSNTTASSPARARSRLTKTKEGFMQPRHFK 1179
 DB 927 -----KGWVHNSID----- 935
 QY 1180 SPVRGQSEQPAQLQPEAQLHVPQAQPOQLQLOQSHKGFLEQEGSPLSLGQSOHDSQ 1239
 DB 936 -----YRFKTHR----- 942
 QY 1240 SAFPILSQTOSHSLSSSVLTPTDSSPGKLDPAQOPPEEPPEPEDEASSPDLQAFWNI 1299
 DB 943 -----KDHNL-----PD-----DYCFRRKKA-----NL 962
 QY 1300 SAOMPCNAAPPPPLAVSSEDQTPSPQOLASSKPMNRPSAANPCSPVQSPSTPLAGLAPKR 1359
 DB 963 GKNASVNAHGPCPALEAVE---TTVPKQ----- 986
 QY 1360 RAGDGMGPMQSPGTGLGQPKRRGRPPSKFFKQMEORYLTQLTAQVPVPMCSGWWIPDPE 1419

```
Db 987 -----GQ-----NLWFLCDSQ 997
Qy 1420 M-LDMLKALPHRGIREKALHKLHNRDPLQEV--CLRPSADPIFEPRQLPAQEGIMS 1476
Db 998 KELDELLSCLHPQIGRESQFKELEKR---YOEITHCNMARKNGLKSCDGNQE-LLN 1053
Qy 1477 WSPKE-----KTYETDLAVLQWVEELRQVIMSDLQIRGTCTSPDSTREDLAYCEHLS 1530
Db 1054 FLRSDLIEVATRLQKGLGYMEGTSEFEARVIS-----LEKXK 1091
Qy 1531 DSQDITWRGPGREGLAGPQRTTNPLDLAVNRLAALAEONVKRYLRLEPLWPTHEVLEKA 1590
Db 1092 DFGEV-----IALQASVIKTFLOGFMAKQK---KRX 1121
Qy 1591 LLSTPNGAPEGTTTEISY-----EITPRIRIWRQTLORCSAAHVCLCLGHLERSIAWEKS 1646
Db 1122 LQSEDSTKSEVDEBEKQVEAKVASALEKWKTAIREAQTFSRMHVLGLMDACIKWDM 1181
Qy 1647 VNKVTCVLCVRKGDNDEFLLCDGDRGCHIYCHRPKMEAVPEGDFWCTVCCLAQVGEFT 1706
Db 1182 AENARCKVCRKKGEDDKLILCECNKAPHFLCLRPALYVEVDGEWQCPAC----- 1231
Qy 1707 QKQFPK--RQKRGSGVSLNFSGDGR-----RRVLLKG 1740
Db 1232 QPPTARRNSRGRNTEESTSGESGDSGESEEBEEDYEVAGLRRLRPRKTI 1291
Qy 1741 RES--PAAGP-RYSEERLSPSKR-----RSLSMNHSHDLTFCEILME 1781
Db 1292 KOSVIPAAAPGRPGKSHPARSRPKDDPVDVLVLTQTKRIS--RRSLELQKCEDILHK 1350
Qy 1782 MESHDAAPFPXEPNPRVLSVGYRIIKNPMDFSTWRERLLRGVTSSEFAADALLVFDN 1841
Db 1351 LVKRFSPFPFEPVTRDEAEYDVIDEHPMDFQIIONKSCGNTSVQFELTDMKQVFAN 1410
Qy 1842 CQTFNEDDSEV 1852
Db 1411 AELYNCRGSHV 1421
```

RESULT 20

```
Q7PRP9 PRELIMINARY; PRT; 1357 AA.
AC Q7PRP9;
DT 01-MAR-2004 (TRENBLrel. 26, Created)
DT 01-MAR-2004 (TRENBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)
DE ENSANGP0000001532 (Fragment).
GN Name=ENSANGG0000001297;
OS Anopheles gambiae str. PEST.
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Anopheles.
OX NCBI_TaxID=180454;
```

(1)

```
SEQUENCE FROM N.A.
RP STRAIN=PEST;
RC Anopheles Genome Sequencing Consortium;
RA Submitted (APR-2003) to the ENBL/GenBank/DBJ databases.
RL -! CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC EMBL; AAB01008847; EAA06826.2; -.
DR HSSP; Q9UIG0; 1F62.
DR GO; GO:0003677; P:DNA binding; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR001487; Bromodomain.
DR InterPro; IPR004022; DDT.
DR InterPro; IPR011011; FYVE PHD_Znf.
DR InterPro; IPR001965; Znf PHD_Znf.
DR Pfam; PF00439; Bromodomain; 1.
DR Pfam; PF00628; PHD; 1.
DR PRINTS; PR00503; BROMODOMAIN.
DR PROSITE; PS00633; BROMODOMAIN_1; 1.
DR PROSITE; PS00014; BROMODOMAIN_2; 1.
```

```
DR PROSITE; PS00827; DDT; 1.
DR PROSITE; PS00016; ZF PHD_2; 1.
FT NON_TER 1357 1357
SQ SEQUENCE 1357 AA; 15305 MW; D42FD563C34F5287 CRC64;
Query Match 5.4%; Score 529; DB 2; Length 1357;
Best Local Similarity 18.7%; Pred. No. 1.2e-13;
Matches 301; Conservative 211; Mismatches 460; Indels 640; Gaps 63;
Qy 558 KRMKQPFVEIK-----YLSRLVSHSVREHPSF--SPRMPVCDPFPEERD----- 599
Db 76 KLLKTFPAAVKGPFLLVASHOTRTSINEMHEDVYGVFKDHLFKGESVDAMDVPRKAFRA 135
Qy 600 -----TPBGLQWVLSABE-----IPSRIOAITCKRGRPRNT--EKAKTKE 638
Db 136 TITNVLVDSLTLPNALYI--QVRAEDDMTPKVTWVLPGNV-----KRDSALTRECKLFL 189
Qy 639 VPVKRGRGRPPKVKITELL-----NKTNRPLKLEAQETLNEEDKAKIAKSKKMRQ 692
Db 190 KQHV EAGPGGQLRIKADSLARFVTGEGWDG-----LVFHGQTPSFQSKRLKQ 238
Qy 693 KVQGECLTTIQOQARNRKQETKSLHKEAKKXAEKEGKTKQSKLKEKVKREKKEK 752
Db 239 NAERGEA-----GGGGAASATALEAKAKQKEEA 267
Qy 753 VMKKEK--EVTKAKPACKADKTLATQRRLEEROKQOMILEMKKPTEDMCLTDHQPDPF 811
Db 268 AQLRLRVEQALAKRAEESKALLAQOVVLAKRYNAVLEDOELP-----DORPLDPV 320
Qy 812 SRVPLGLTPSGAFSDCLTIVEFLHSFGKVLGDFPAKDVP---SLGVLQEGLLCGDSLGE 868
Db 321 RPVRPL--IAARHFSSFFILEYLSNFPADLLAI--RSKFPGLTTHLLERALILR--EVNGP 376
Qy 869 VQDLLVRLKKA---ALHDPGFPSP-----YCQSLKILG 897
Db 377 LSDIFQVLLSAIFTQLEENEESESVRYERIEGLAQKRTVPQVRRARDAALWCE--KHYS 434
Qy 898 EKYS EIPLTRDNVSEILRCFLMAYGV-----XPAL 927
Db 435 TKLSEMPIDSMVTSELRLHFLASGALVVEEARHRYQNRGGYSGDDPGLRLVRDPHI 494
Qy 928 CDRLRTQPFQAPPPQKAAVLAFFPVHEL---NGSTLIINE-IDKTLSEMSYRKNKIVE 983
Db 495 LRALNWSYVYQLPVGDIILCCLIHQLLTYSVGRDLVEERVEKARTARTSYQANRW--- 551
Qy 984 GLRLRLKTVLA---KRTGRSEVEMGRPEECGLRRRRSRIMEETSMEEEEEESIAAVPG 1040
Db 552 --AQRRITVTVKAGSLKNAARDMK-----RELAGFEGE-----LAAREE 587
Qy 1041 RRGRRDGEVDATASSIPELEROIEKLSKROLFFPKLLHSSQMLR-----AVSLGODRY 1094
Db 588 YRKLRLRLDEQCA-----RIEADAQRL---KVLQAESERLKEDEFFDYQIYLGTDR 637
Qy 1095 RRRYVLPYLAGIFVEGTEGNLVPEEVIKKTETSLKVAHASLNPALEFMSKMLAGSNNT 1154
Db 638 YRNYWLPESLPGLFVEHDR----- 656
Qy 1155 ASSPARARSRLPKTKPGFMOPRHPKSPVRGQDSEPOQAQLOPEAQHLVPAQPQOLQLOL 1214
Db 657 -TVAGRCILDRPTPHIPGL-----AACAPD----- 679
Qy 1215 QSHKGFLEQSGSPLSLGCSQHDLSQAFSLWSLQTSQSHSLSSSVLTTPDSSPGKLDPA 1274
Db 680 -QRKKYITRA-----IMRCATDGVVALDDAKAFA-DAN-----AV 713
Qy 1275 SQPPEEPDEAESSPDLQAFNFIQAQMPNCNAAPTPLAVSEDOPTSPQOLASSKPMN 1334
Db 714 SSSSTGPAVVEE-----LLMCTAQP-----DDCPVHTEQH----- 744
Qy 1335 RPSAANPCSPVQPSSTPLAGLAPKRRAGDQEMQPSPTGLQPKRGRGPPSPKFFQMBQR 1394
Db 745 -----PGTV----- 748
```


DR InterPro; IPR001487; Bromodomain.
DR InterPro; IPR004022; DDT.
DR InterPro; IPR001965; ZnF PHD.
DR Pfam; PF00439; Bromodomain; 1.
DR Pfam; PF00628; PHD; 1.
DR PRINTS; PR00503; BROMODOMAIN.
DR SMART; SMO0297; BRMO; 1.
DR SMART; SMO0571; DDT; 1.
DR SMART; SMO0249; PHD; 1.
DR PROSITE; PS00633; BROMODOMAIN_1; FALSE_NEG.
DR PROSITE; PS00014; BROMODOMAIN_2; 1.
DR PROSITE; PS00827; DDT; 1.
DR PROSITE; PS01359; ZF PHD 1; 1.
DR PROSITE; PS00016; ZF PHD 2; 1.
KW 3D-structure; Alternative splicing; Bromodomain; Coiled coil;
KW Nuclear protein; Transcription regulation; Williams-Beuren syndrome;
FT Zinc-finger.
FT DOMAIN 20 126 WAC.
FT DOMAIN 604 668 DDT.
FT ZN_FING 1184 1234 PHD-type.
FT DOMAIN 1356 1426 Bromodomain.
FT DOMAIN 306 578 Lys-rich.
FT DOMAIN 533 586 Coiled coil (Potential).
FT DOMAIN 768 814 Coiled coil (Potential).
FT DOMAIN 850 893 Coiled coil (Potential).
FT DOMAIN 1245 1283 Coiled coil (Potential).
FT DOMAIN 1261 1273 Poly-Glu.
FT VARSPIC 660 663 Missing (in isoform 2).
FT CONFLICT 14 14 /FTid=VSP_000552.
FT CONFLICT 22 22 K -> N (in Ref. 3).
FT CONFLICT 136 136 L -> F (in Ref. 3).
FT CONFLICT 191 191 K -> E (in Ref. 1).
FT CONFLICT 298 298 N -> D (in Ref. 4).
FT CONFLICT 823 823 Y -> V (in Ref. 4).
FT CONFLICT 1191 1191 E -> R (in Ref. 3).
FT CONFLICT 1354 1354 R -> P (in Ref. 3).
FT CONFLICT 1438 1438 K -> M (in Ref. 2).
FT CONFLICT 1438 1438 A -> V (in Ref. 3).
SQ SEQUENCE 1483 AA; 170902 MW; 0CC146FEBB954261 CRC64;

Query Match 5.4%; Score 528.5; DB 1; Length 1483;
Best Local Similarity 19.5%; Pred. No. 1.4e-13;
Matches 265; Conservative 179; Mismatches 439; Indels 479; Gaps 42;

QY 619 QAITKGRPR--NTEKAKTEVPKVRGRPRPKVKITEL-----LNKTDNRPLKCL 669
DB 415 QKSTGNSPKKGLKTPKTKQMTLLDMAGTQ---KMTAPRNSGGTPTRTSKPKHKL 471
QY 670 EA-----QETLNEEDKAKIAKSKKMRQKVRQGECLTTIQQARNKRKQETSKLKH 720
DB 472 PPAALHLIAVYKENKDRDKKSALSCVISKTARLLSSEDRARLPPEELRSLVQRYELLEH 531
QY 721 KE--AKKKSXAEGKGTQKQKKEKVRKEKVKRKEKEVTKAKPACRADKTLATOR 778
DB 532 KKRWASMEEQKVELKKREELKKLKEKAKER---REKE----- 569
QY 779 RLSEKQKQMLEEMKPTEDMCLTDHQPDPFSRVPLGT-LPSGAFSDCLTIVEFLHSP 837
DB 570 MLERLEKQ-----KRYEDQELTG-KNLPAPFLVDTPEGLPNTLFGDVAMVVEFLSCY 620
QY 838 GKVLGFDPAKDVPSLVLQEGLLCGDGLSGVQDLLVRLLLKAAALHDPGFPSPYCSOLKTLG 897
DB 621 SGLLPDAQYITAVS-LMEALSADKGGFLYLNRVLLVILLQTLQDETAEDYGE-----LG 675
QY 898 EKVSIEPLTRDNVSEILRCFLMAYG-----VXPALCDRLRTQ 934
DB 676 MKLSEIPLTHSVSELVRLCLRRSDVQSESGSDTDNNDKDSAAFEDNEVQDEFLEKLETS 735
QY 935 PFQAPQPOKAAVLAFPVHELNGSTLIINEIDKTLKESMSYRKNKWIIVEGLRLRK----- 990
DB 736 EFFELTSEEKLOIITALCHRI-----LMTYSQDHMETRQQMSABLW--KERLVLKEND 789
QY 991 TVLAKRTGRSEVEM-----GRPECLGRRRSRIMEETSGMEEBEEBEEBESIAAVPGR-- 1042

DB 790 KKRAEKQKRKMEAKNKENGKVENGLGKTDRKKEIVKFEPOVDTEAEDMISAVKSRLLA 849
QY 1043 --GRDGEVDATASSIPELEROIEKLSKR-----QLFFRKKLLHSSQMLRAVSLGDYR 1094
DB 850 IQAKKERBEIOEREMKV-KLERQABEEERIRKHAAAEKAFQEGIAKAKLVWRRTPIGTDNR 908
QY 1095 RRRYWL-PLYAGIFVETGTEGNLVPBVIKKTETSLKVAHAASLNPALFSMKMELAGSNT 1153
DB 909 HRYWLFSDVPGLFIE-----KGWVHDSID----- 934
QY 1154 TASSAPARARSLKTKPGFMQPRHKS-PVRGDSQEQQAOLPEAOHLVPAQPOPOLQL 1212
DB 935 -----YRFNHKCHTVDGSDYCPRSK----- 957
QY 1213 QLQSHKGFLEQSGSPLSGOSQHDLSQSAFLSWLSQTQSHSSLLSSSVLTDPDSSPGKLDP 1272
DB 958 -----KANLGN-----ASMTQHGTAATEVAVEITTPKOG----- 987
QY 1273 APSQPPEPEPDEAESPDLOAFWNISQAQPCNAAPTPLAVSEDOQTPSPQQLASSKP 1332
DB 988 -----QNLWFLCDSQ----- 997
QY 1333 MNRPSAANPCSPVQFSSTPLAGLAPKRAGDPMGEMPOSTGLGQPKRRGRPPSKFFKQME 1392
DB 998 ----- 997
QY 1393 QRYLTQLTQAPVPPMCSGWWITPDPEMLDAMLKALHPRGIREKALHKLHNK-HRDFLOE 1451
DB 998 -----KELDELLNCLHPQIGRESQLKREKRYQDIHS 1031
QY 1452 VCLRPSADPIFEPRQLPAPFOEGIMSWSPKE-----KTYETDLAVLQWVELEQRVMSD 1505
DB 1032 IHL--ARKPNLGLKSCDGNQE-LLNFLRSLDILIEVATRLQKGLGYVETSEFEARVLS-- 1086
QY 1506 LQIRGWTCPSPDSTREDLAYCEHLSDSQEDITWRGPGREGLAPQKRTNPLDLAVMLAA 1565
DB 1087 -----LEKLKDRGECV-----IA 1099
QY 1566 LEQNVKRYLREPLWP-----THEVLEKALLSTPNGAPEGTTTTSIEVEIT 1611
DB 1100 LQASVKKFLQGFMAPQKRRKLQSEDSAKTEEVDEEKGWVE-----EAKVA 1146
QY 1612 PRIRIWOTLQRCRSAHVCLCLGLHRSIAWKSVNKVTCLVCRKGNDEFLLLCGCD 1671
DB 1147 SALEKWTATIREAQTFSRMVLLGLMDACIKWDSAEARNACKVCRKKGEDDKLILDECN 1206
QY 1672 RGCHYCHRPKMEAVPEGDMFCTVC-----LAQQVEGEFTQKPGPK 1713
DB 1207 KAFHLFCLRPALYEVPDGEMQCPACQPATARRNSRGRNYTEESASEDSEDEEBEE 1266
QY 1714 RGQVKSGYSLNFBEGDGRRRVLLKGRES---PAA--GPR-----YSEERLSPK---- 1759
DB 1267 EEEBEEEDYEV--AGURLPRKTIIRGKSHVIPPAAARSGRRPGKKPHSTRSQKAPPVD 1323
QY 1760 -----RRRLSMRNHSDLTFCETILMEMESHDAAPFEPVNPRLVSGYRRIKNP 1810
DB 1324 DAEDVELVLTQKSSRRQSLQLEKCEILHKIVKYSFWPFRPVTDEADYDVTIHP 1383
QY 1811 MDFSMTRELLRGGYTSSSEFAADALLVFDNCQTFNEDDSEV 1852
DB 1384 MDFQTVQNKCSGYSRVSQEFLTDMKQVFTNAEVYNCRGSHV 1425

RESULT 22
QYV9T4
ID QYV9T4 PRELIMINARY; PRT; 1476 AA.
AC QYV9T4;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)
DE CG1966-PA.
GN Name=Acf1; ORFNames=CG1966;


```
Db 419 TYLDRAAQTQBPYWSMAQAASHLYAKRHFSKVNELPLDALTLSEVLRLHLLSGAFVN 478
Qy 924 -----XPALCDRLRTQPPQAQPPQQAVALAF-----PVHELNGSLI 961
Db 479 EKAERWRVWYNGYSSKEDPGLURL-----EHSILRLKNHVSYQLKFXDIM 527
Qy 962 -----INEIDKTLSEMSYRKYN-KWIVEGRLRLKTLVAKRTGRSEVEM 1004
Db 528 LLIRCLMSQIMWYSGTINLIERMEQTAKARQDLRALVVGENKL-----NAVEI 577
Qy 1005 GRPECILGRRSRIMETSQWEEEBEESIAAVPGRRGRDGEVDATPASSIPELERQIE 1064
Db 578 NR-----KLTQMHLLEVNGVEPE-----KREALVEKLKKSIAELHAQSD 617
Qy 1065 KLSKROLFFRKLIHSSOMLRAVSLGQDRYRRRYWVLPYLAGIFVEGTEGNLVEEVKK 1124
Db 618 QOHRKHLOMLK-LHSQLFNLFVLGMDCRYKRYVLESMPGIFVE-----H 663
Qy 1125 ETDLSIKVAHASLNPALFMSKMLAGSNNTTASSPARARSRLKTKPGFMQPRHFKSPVRG 1184
Db 664 SPDSL-----DTCLEQPIYTKS----- 680
Qy 1185 QDSQPQALOPEAQHLVPAQPOLQLOLQSHKGFLEQGSPLSQOSHDLSQSAPLS 1244
Db 681 -----QIEIRQQSAL-----PKNRKDLRVLLKLYGDDEKK--TKKAKHSLN----- 722
Qy 1245 WLSQTQSHSLLSSVLTPDSPSGKLDPAQPPPEEPDEAESPDLOAFWNISQAOMP 1304
Db 723 --KENQEHR-----LNGSAE-- 735
Qy 1305 CNAAPTPLVASEDOPTSPSPOOLASSKPMRPSAANPCSPVQFSSTPLAGLAPKRRAGDP 1364
Db 736 -----PMDVESDPE-----APTHFELLMCSG--DRSC-- 762
Qy 1365 GEMPQSTGLQOPKRRGRPPSKFFQMBQRYLTQTAQVPVPEMCSGWWIIPDEMLDAM 1424
Db 763 -----IVHDSRNGQ-----RQR-----WAYIYKAEEDL 787
Qy 1425 LKALHPRGIRKALKHKLKHDRDFOE-----VCLRPSADPIPEPQLPAFOQGISWS 1478
Db 788 IKALNPNGRLREYELLQELSVLRSIEQHAQKCPVDLLSLENETWRKFKPMAAME-----S 841
Qy 1479 PKEKTY-----ETD-----LAVLQWVEELEQRVIMSDL-----QJRGWTCPSPD 1517
Db 842 ETNRKYGEANFGLPNGTDLNEVRLHLVDRIIQENDIYTGDLGRLKVKDMEK----- 895
Qy 1518 STRDL-----AYC-----EHLSDSQEDIT-----WRGPR 1543
Db 896 --RSDLLGNYDAQCKLQWPGGKLEDEAGSDNESHETHEDDGALLGKYARKPYRDPGM 953
Qy 1544 EGLAPQKT-----TN--PLDLAVMLAA-----LEQNVKRYLREPL-----W 1580
Db 954 Y-LAASADTKPLPDSDDEOQHTNAVSIPIAVHNNASALLQVEAIGKRLFKPEYGMKKW 1012
Qy 1581 PTHEVLEKALLSPNGAPEGTTTIEISYEIPTRIRWQTLQRCRKAHVCLCLGHLERS 1640
Db 1013 DPKEALKLADS-----RLHQWEVSLMESTSPAQVFLHNLHLD 1053
Qy 1641 IAWKSVNKTVCVCRKGDNDDEFLLDCGDRGCHIYCHRPKMAVPBGDWFCTVCL--- 1697
Db 1054 IQWRSTNKSLSCKVCRGSDPEKMLLDCENAGTHMFLCKPLRSVPPGNWYCNDCVKS 1113
Qy 1698 ----- 1697
Db 1114 GLSNGQNEKDKQATKKRKFIVEEDDEATDEBEKKDDMTDEDAEHENEKHDEDE 1173
Qy 1698 -----AQQVEGEFTQKPGPKRGOKRKS-----GYSLNFSEG 1729
Db 1174 DDESVTSTPSSRVNGRLIRPRTRPTSRRLTSKEIEBHAQEDVDSGDVSDASLTAGED 1233
Qy 1730 -----DG-----RR 1733
Db 1234 TIEDESDEKVCQKCFYDGGBIKVCQCLRFFHLECVHLKRPPTDFTVCKTCKPMPQPRR 1293
```

```
Qy 1734 RRVLLKG-----RESPAAGPRYGEERLSPSK-----RRRLSM 1765
Db 1294 RHSNMNGDHRDDEEPPAKPRNSRLRLSIDKTARPSNGNNNNNNSSVNNHRRSGRR 1353
Qy 1766 RNHSDLTFCBI--ILMESHDAAPFPXBPVNPRLVSGYRRRIIKNPMDFSTMRELRLLRG 1823
Db 1354 TNEHPLNSAALYDLLSQIMKHAAPFLRPVLVTSEVPDYHQIITKTPMDLAKIKSKLNMG 1413
Qy 1824 GYTSSSEFADALLVFNQCOTFNEDDSEVGKAGHIMRRFFESR 1866
Db 1414 AYQLNEELLSDIQLVFRNCGLYNVEGNEIYDAGCQLERFVIDR 1456
RESULT 23
QY0Y0W1
ID AC QY0Y0W1 PRELIMINARY; PRT; 1476 AA.
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE ATP-dependent chromatin assembly factor large subunit.
GN Name=Acf1;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99315627; PubMed=10385622;
RA Ito T., Levenstein M.E., Fyodorov D.V., Kutach A.K., Kobayaashi R.,
RA Kadonaga J.I.;
RT "Acf consists of two subunits, Acf1 and ISWI, that function
RT cooperatively in the ATP-dependent catalysis of chromatin assembly.";
RL Genes Dev. 13:1529-1539(1999).
DR EMBL; AF148962; AAD38952.1; --
DR HSP; Q9UIG0; 1562
DR Flybase; FBgn0027620; Acf1.
DR GO; GO:0016590; C:ACF complex; IDA.
DR GO; GO:0008623; C:chromatin accessibility complex; IPI.
DR GO; GO:0003677; F:DNA binding; TAS.
DR GO; GO:0042766; P:nucleosome mobilization; IDA.
DR GO; GO:0016584; P:nucleosome spacing; IDA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IDA.
DR InterPro; IPR001487; Bromodomain.
DR InterPro; IPR004022; DDT.
DR InterPro; IPR001965; ZnF_PHD.
DR Pfam; PF00439; Bromodomain; 1.
DR Pfam; PF00628; PHD; 2.
DR PRINTS; PRO0503; BROMODOMAIN.
DR SMART; SM00297; BROMO; 1.
DR SMART; SM00571; DDT; 1.
DR SMART; SM00249; PHD; 2.
DR PROSITE; PS00114; BROMODOMAIN_2; 1.
DR PROSITE; PS00827; DDT; 1.
DR PROSITE; PS01359; ZF_PHD_1; 2.
DR PROSITE; PS00016; ZF_PHD_2; 1.
SQ SEQUENCE 1476 AA; 170457 MW; C18A6CEBF57E66B3 CRC64;
```

```
Query Match 5.4%; Score 526.5; DB 2; Length 1476;
Best Local Similarity 19.8%; Pred. No. 1.7e-13;
Matches 341; Conservative 202; Mismatches 487; Indels 693; Gaps 66;
Qy 530 QHGWRREVRIKKGSHRWQGETWYYPGCKRMKQPEVIKYLNRNLVHSVRREHFSFSPRM 589
Db 41 EHYFRHVMVI--NSTVQCEA----TGKENTVTEAVK--SERAARK-KMEQKQSLRA 90
Qy 590 PVGDFEERTPEGLQWVQLSABEIPSRIOAITKRCRPR-----NTEKAKTKEVP 640
Db 91 PV-----LLVVEHAQSAVNTLNNIVAKFLRKRYFTGEEVSVQAKKATYVL 138
Qy 641 KVKRGGRPPKVKITELLNKTDN-----RPLK-----KLEAQ-ETLINE 677
```

Db 139 GVLKDNWPE--PLNGIYEDTDLNVLRLPNKGDPSAELDLFPRLQRLSRMEFNLENLSM 196
QY 678 EDKAKIAKSKKWKORQVORGECLTTIQGQ-----ARNKRQETKSLKHEAKKS--- 727
Db 197 FIKSNVSRVDGLLRPKPEAYKYVTPNGVNFSTIFIGKMPRYSAPAK-INKPDGKQSTLN 255
QY 728 -----XAEKEKTKQEKULK-----EKVREKKEKVKMEKESEVTKAKPACAKADKYL 774
Db 256 KYIVAGEATAAKSAKAKSAKSAKSAKSAKSAKSAKSAKSAKSAKSAKSAKSAKSAK 304
QY 775 ATQRLERQKQOMILEMKKTEGMDCLTDHQPDPFSRVPGLTLPSPGAFSDCLTIVEFL 834
Db 305 --KAQLIERVENECNL--LQKTDLLERTDQKVLPRYQIVTL-LPEHLLGDAFWMBFPM 359
QY 835 HSF-GKVLGDFPAKDVPSLGLVQEGLLCOGDSLGEVQDILLVRL 877
Db 360 HYYTGLLSGIEVFRQNLSEYENTRALTAR-ETAGPLSDILLVLLGTVPDLQKEEBEBCAV 418
QY 878 ----KAALHDPGFPSCYQSLKI-----LGEKVEIPLTRDNVSEILRCFLMAYGV--- 923
Db 419 TYLDRAAQTQEPYWSMAQAASHLYAKRHFSFKVNELPLDALTLSEVLRLLHLLGSAFVN 478
QY 924 -----XPALCDRLRTOPFOAQPPQKAAVLAF-----PVHELNGSTLI 961
Db 479 EKAERWVRMYRNGYSKEDPGLRL-----EHSHTLRILKNHSVYQLKFKDIM 527
QY 962 -----INEIDKTLSSMSYRKN-KWIVEGRULRLKTLVAKRTGRSEVEM 1004
Db 528 LLIRCLMSQIMTYSGTINLIERMEQTAKARQDLRALVVGKRL-----AAVEI 577
QY 1005 GRPECLGRRRSRRIEETSGMEEBEESIAAVFGRGRDRGEVDATASSIPELERQIE 1064
Db 578 NR-----KCLTQMHLEVNGVEPE-----KREALVEKLKKSIAELHAQSD 617
QY 1065 KLSKQLFRKLLHSSQMLRAVSLGQDRYRYYVLPYLAGIFVBTGEGNLVPEVIKK 1124
Db 618 QOHRKHELQMLK-LHSQLFNVLVGLMDRCYKYYLVESMPGIFVE-----H 663
QY 1125 ETDLSLVAHAASLNPALFSMKMELAGSNTTASSPARAKSRPLKTPGFMQPRHFKSPVRG 1184
Db 664 SPDSL-----DTCLEQPTNKS----- 680
QY 1185 QDSEQPAQLOPEAQLHPAQPQLOLQSHKGFLEQGSPLSGOSHDLSOAFLS 1244
Db 681 -----QIEIROQSAL-----PKNRKDLRVLLKLYGDDEKK--TKKAKHSLN----- 722
QY 1245 WLSQTQSHSLSSVLTPDSFGKLDPAPOSPEPEPEDEAESPDLOAFWNFISAOMP 1304
Db 723 --KENQEHR--LNGSA-----EPMQVESN----- 742
QY 1305 CNAATPPPLAVSEDTPTSPQQLASSKPMNRPSAANPCSPVQFSSTPLAGLAPKRRAGDP 1364
Db 743 -----SPEATHFELLMCSG--DKRSC--- 762
QY 1365 GEMPOSPTGLQPKRGRPPSKFFKQMEQRYLTQLTQAPVPPPEMCSGWWIWDPPMLDAM 1424
Db 763 -----IVHDSRNGQ-----RQR-----WAIYKAESEIDEL 787
QY 1425 LKALHPRGIREKALHKLHNRDFLOE-----VCLRPSADIFEPRLPAPQEGIMSWS 1478
Db 788 IKALNPNGLREVELLQELSLVSLRLEQHAQKTPVDLLSLENETMRKFKWAME-----S 841
QY 1479 PKEITY-----ETD-----LAVLQWVELEQRVIMSL-----QIRGWTCPSPD 1517
Db 842 ETNRKYGANFGLPNGTDLNEMVRLHLDVRIIFQENDIYTGDLRLKVKOMEKW----- 895
QY 1518 STREDL-----AYC-----EHLSDSQEDIT-----WRGGR 1543
Db 896 --RSDLLGNYDAQCKLQWPGCKLEDEAGSDNESHETHEDDQALLGKYARKPYRDFGM 953
QY 1544 EGLAPQRTK-----TNP--LDLAVMLAA-----LEQNVKRYRLREPL-----W 1580

Db 954 Y-LAASADTKPLPDSDDDEQHTNAVLPIAVHNNASALLQVEQAIGKRFLKBPYGMKKM 1012
QY 1581 PTHEVVLKALLSTENGAPEGTTTTEISYEITPRIRIWRQTLQRCRSAHVCLCLGHLERS 1640
Db 1013 DPQKQALACDS-----RLHQWESVLMESTSPAQVFLHLNLHDC 1053
QY 1641 IAWKSVNVKTVCLVCRKGNDBFLLCDGCRGCHYIYCHRPQMEAVPBGDFWCTVCL----- 1697
Db 1054 IQMRSTNKSCLKVCRGSDPEKMLLCECNAGTHMFLCKPKLRSVPPGNYCNDCKVSL 1113
QY 1698 ----- 1697
Db 1114 GLSHGQNEKDKQATKKRKFIVEEDDEATDEBEEKKDDMTDEDABEHENEKDEDEVE 1173
QY 1698 -----AQQVEGEFTQKPGFPKRGQKRS-----GYSLNFSG 1729
Db 1174 DDESVTSTPSSRVNGRILRRPRTPTSRRLTSKEIEHAQEDVSDGVDSDASLTAGD 1233
QY 1730 -----DG-----RR 1733
Db 1234 TIEDESDEKVCQKCFYDGEIKVCQRLFFHLECVHLKRPRTDFVCKTKPMPQPRR 1293
QY 1734 RVLVKG-----RESPAAGPRYSEERLSPSK-----RRLSM 1765
Db 1294 RSHNMGDHDDEDEPKAKRPRNSURLSIDKTARPSNGNNNNNNNNSSVNNNNHRRSGR 1353
QY 1766 RNHSHDLTFCEI--ILMEMESHDAAPPEKPEVNPRLVSGYRRIIKNPMDFSTMRRLRG 1823
Db 1354 TNEHPLNSAALYDLLEQIMKHKAAPFLRPVLTSEVPDYHQIITKPMDLAKIKSLANG 1413
QY 1824 GYTSEEFADALLVDFNCQTFNEDDSVKGAGHIMRRFFBSR 1866
Db 1414 AYQLNEELLSDTQLVFRNCGLYNVEGIEYDAGCOLERFVIDR 1456

RESULT 24
Q9NG24
ID Q9NG24 PRELIMINARY; PRT; 1476 AA.
AC Q9NG24;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE Chromatin accessibility complex (CHRCAC).
GN Name=Acfl1;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21340270; PubMed=11447119; DOI=10.1093/emboj/20.14.3781;
RA Eberharter A., Ferrari S., Langst G., Straub T., Imhof A.,
RA Varga-Weisz P., Wilim M., Becker P.B.;
RT "Acfl1, the largest subunit of CHRCAC, regulates ISWI-induced nucleosome
remodeling";
RL EMBO J. 20:3781-3788(2001).
DR EMBL; AJ238397; CAB86669.1; --
DR HSP; Q9UG0; I562.
DR Flybase; FBgn0027620; Acfl1.
DR GO; GO:0016590; C:ACF complex; IDA.
DR GO; GO:0008623; C:chromatin accessibility complex; IPI.
DR GO; GO:0003677; F:DNA binding; TAS.
DR GO; GO:0042766; P:nucleosome mobilization; IDA.
DR GO; GO:0016584; P:nucleosome spacing; IDA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IDA.
DR InterPro; IPR001487; Bromodomain.
DR InterPro; IPR004022; DDT.
DR InterPro; IPR001965; Znf PHD.
DR Pfam; PF00439; Bromodomain; 1.
DR Pfam; PF00628; PHD; 2.
DR PRINTS; PR00503; BROMODOMAIN.
DR SMART; SM00297; BROMO; 1.

RT "ISWI remodeling complexes in Xenopus egg extracts: identification as
RT major chromosomal components that are regulated by INCENP-aurora B."
RL Mol. Biol. Cell 13:25-39(2002).
DR EMBL; AF412333; AAL60161.1; --
DR HSSP; Q9UIG0; 1F62.
DR GO; GO:0000151; C:ubiquitin ligase complex; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0004842; F:ubiquitin-protein ligase activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0016567; P:protein ubiquitination; IEA.
DR Pfam; PF00439; P:regulation of transcription, DNA-dependent; IEA.
DR Pfam; PF00628; PHD; 1.
DR PRINTS; PRO0503; BROMODOMAIN.
DR SMART; SM00297; BROMO; 1.
DR SMART; SM00249; PHD; 1.
DR SMART; SM00184; RING; 1.
DR PROSITE; PS00014; BROMODOMAIN_2; 1.
DR PROSITE; PS00827; DDT; 1.
DR PROSITE; PS01359; ZF_PHD_1; 1.
DR PROSITE; PS00016; ZF_PHD_2; 1.
DR PROSITE; PS00089; ZF_RING_2; 1.
FT NON_TER 1
SQ SEQUENCE 1079 AA; 124503 MW; 2AF4689FFE9E83DF CRC64;

Query Match 5.3%; Score 519; DB 2; Length 1079;
Best Local Similarity 19.3%; Pred. No. 2.4e-13;
Matches 265; Conservative 185; Mismatches 406; Indels 518; Gaps 47;

QY 619 QAITCKGRPRNTEKA-----KTEKVPKVKR-----GGRPPKVKITELLKNTNR 664
DB 29 QKVSCKTRSPKAVKSPKLMQMTLLDMAKTPKSVRAQKGSSTP-----RSSSK 78

QY 665 PLKLEA-----QETLNEEDKAKTAKSKKMKQKQVQGECLTTIQGARNRKKOET 715
DB 79 PHKYLPPAALHLISYFRDNKNEDR--KSALS--ALISKVAR--MLTA-----EDRKLPL 127

QY 716 KSLKHKEAKKXSAB--KEKGKTKQKELKEKVKREKKEKVKMKKEBEVTKAKPACKADKYL 774
DB 128 DDQLQVLQKRYDLLEHRKQWAVMTQEEREYMKKREALKVRIKEKARERK----- 178

QY 775 ATQRLEROKOIMLEEMKPTEDMCLTDHOPDPFSRVPCLT--LPSGAFSDCLTIVRF 833
DB 179 ---QKEREERLEKQ-----KRYEODEITG--KSLPTFKLVDTPGELPNALFGDVAWVIEF 228

QY 834 LHSFKVLGFDPAKDPVSLGVQLQEGLLCGDSLGEVQDILLVRLKAAALHDPGFPGSYCOSL 893
DB 229 LSGYSDDLPLDPAQYPTAVS--LMEALAAEKGGFMYLNRVLVLLQTLQDEITAEDYGE-- 285

QY 894 KILGKVKSEIPLTRDNVSEILRCFLM-----AYG-----VXPAL 927
DB 286 --LGMKLSIPLVLSASSELVRLCLRKSDSPAGESDSTEKGDEDSVGSVAVYQDDDEVEY 343

QY 928 CDRLTQTPQAPPOOKAAVLAPFPHELNGSLIINEIDKTLNEMSSYRKNKWIIV--EG 984
DB 344 LEKLTSEFFELTTEKHLILAAALCHRLIMTYVSQDHDVDAKQORSGEIWLKERLALKEEN 403

QY 985 RLRLKTLVAKRTGTSEVMGRPEPCLCGRSSRIMEETSGMEEBEEESIAAVPGR-- 1042
DB 404 DKKR-----AEKQKEKEQTVKPEVQAAKIVKQEKINTQDNDADMISAVKSRRLQ 458

QY 1043 ---GRDGEVDATASSIPELERQI--EKLSSKQLPFRKKLLH-----SSQMLRAVSLGQDR 1093
DB 459 AMQAKKEKEHEKLTG--ERIERETEERSRKQKASAEKAFEGIAKAKVLRLRRSPGLGTR 517

QY 1094 YRRRYTWL--PYLAGIFVEGTENLVPEVIVKKTDSLKVAHAASLNALPSMKMELAGSN 1152
DB 518 NHNRYWLFSEVPGIYIE-----KGWVHDSIN----- 544

QY 1153 TTASSPARSRPLTKPGFMQPRHFKSPVRGQDSEQPAQLOQPEAQLHVPAPQPOQLQL 1212
DB 545 -----YRFPSEKQDSEQDAE----- 561

QY 1213 QLQSHKGFLQEGSPSLSGQSHDLQSAPLWSLQTSQSHSLSSSVLTPDSSPGKLDLP 1272
DB 562 -----SEDA-----NSSIGCPDSS----- 575

QY 1273 APSQPPEPEDEABESSDPLQAFWFNISAQMPCNAAPTPLAVSDDQPTFSPQQLASSKP 1332
DB 576 ----- 575

QY 1333 MNRPSAANPCSPVQVSTPLAGLAPKRRAGDPGEMPOSPTGLGQPKRRGRPPSKFKQME 1392
DB 576 -----TQRE 579

QY 1393 QRYLTQLTQAPVPPPEMCSGWMIIPPEM--LDAMLKALHPRGIREKALHKHL--NKHRDFLQ 1450
DB 580 EKX-----AETVPKQGGQNLMPFLCTQKELDELLSLHPQGFRESQKLERLQRYQDINH 634

QY 1451 EVCLRPSADPIPEPRQLPAFQEGIMSWSPKE-----KTVETDLAVLQWVE---ELEQRYI 1502
DB 635 SIHL-----ARKQNLGLKTCGQQLNLFRLSRDIIIEVATRLQ 671

QY 1503 MSDLIQIRGWTCPSPDSTRE---DLAYCEHLSDSQSDITWRGFGREGLAPQRKTTNPLOIA 1559
DB 672 KGGI---CYL---DDTTEFEAKVRTFENLXDFGECIVF-----LQAA 707

QY 1560 VMR-----LAALQNVKRYRLREPLWPTHEVLEKALLSTNGAPGEGTTTSEIYSITPRI 1614
DB 708 VIKKFLQGFMAPKQKRRKHQSEAAKAEQDEEKKM-----AEEA-----KVASAV 754

QY 1615 RIWQTLQRCSSAAHVCLICLGHLESIKAWKSVNKTCLVCRKGDNDFFLLCDGCDRGC 1674
DB 755 EKWKVAIKDAQTFRSMHVLLGLMDACIKWDMSSENARCKVKKGEDDKLLICDCNCAF 814

QY 1675 HIYCHRPMEAVPEGDFWCTVCLAQVQEGEFTQKPPFPKQKRSYSLNFSE----- 1728
DB 815 HLFCLRPVLFVNPDEGLWCPAC-----QPATARSSGRNVYAEASTQDE 858

QY 1729 -----GDGRRRRVLLKGR-----ESPAA 1746
DB 859 DEEBEESSEBEEBESDEEBEEOEMMQRLRSRAAKGKGPRTTRGRPPKNNTHSRVS 918

QY 1747 GPRYSEE-----RLSPSKRRLLSMRNNHSDLTFCETILMEMSHDAAMPKPEPVN 1796
DB 919 RQRYVEDTEADVEMVROSKEPTSRQ-----NOEFQCEETLAKLITKRFSPWPREPN 972

QY 1797 PRLVSGYRIIKPMWDFSTMRELLRGGYTSSEFAADALLVDFNDCOTFNEDDS 1850
DB 973 ADEIEDYTKVVTTPMDFTMQSKCSGYSQYTVQEFNLNDLKLVPFGNTELYEAGS 1026

RESULT 26
Q9NSL9
ID Q9NSL9 PRELIMINARY; PRT; 1427 AA.
AC Q9NSL9;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Hypothetical protein H20J04.2.
GN Name=H20J04.2; ORFNames=H20J04.2;
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=99069613; PubMed=9851916;
RG WormBase Consortium;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN (2)
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;


```
Qy 1587 LEKALLSTPNGAPEGTTTETISVEITPRIRIWRQTLQRCRSHAAHVCLCLGHLERSIAWEKS 1646
Db 973 --KRCRDRN-----TVAMVFRGOKRLDWEGLMRATSYSLFLHYHLYLDYWSRS 1025
Qy 1647 VNKVTCVLCVRKGDNDFFLLCGDGRGCHYIYCHRPKMEAVPEGDFWCTVCVLAQVGEFT 1706
Db 1026 IERAVCMVCRKGDANLTLLDCNCRACHMYCLPKLVKVPBGDFWCFKWCPSDDAGIVS 1085
Qy 1707 QK 1708
Db 1086 RK 1087

RESULT 28
Q6P1D9 PRELIMINARY; PRT; 892 AA.
ID Q6P1D9;
AC Q6P1D9;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE BC065123 protein (Fragment).
GN Name=BC065123;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Brain;
RX MEDLINE=22389257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettner M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smallos D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Brain;
RA Strausberg R.;
RL Submitted (JAN-2004) to the EMBL/GenBank/DBDJ databases.
DR EMBL; BC065123; AAH65123.1; -.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR001487; Bromodomain.
DR InterPro; IPR011011; FYVE PHD_Znf.
DR InterPro; IPR001965; Znf PHD_1.
DR Pfam; PF00439; Bromodomain; 1.
DR Pfam; PF00628; PHD; 1.
DR PRINTS; PR00503; BROMO; 1.
DR SMART; SM00297; BROMO; 1.
DR SMART; SM00249; PHD; 1.
DR PROSITE; PS00633; BROMODOMAIN_1; 1.
DR PROSITE; PS50014; BROMODOMAIN_2; 1.
DR PROSITE; PS01359; ZF PHD_1; 1.
DR PROSITE; PS50016; ZF PHD_2; 1.
DR NON_TER 1
SQ SEQUENCE 892 AA; 101830 MW; 95A6C798F37AA93F CRC64;
```

```
Query Match 4.8%; Score 469.5; DB 2; Length 892;
Best Local Similarity 20.8%; Pred. No. 2e-11;
Matches 230; Conservative 114; Mismatches 255; Indels 507; Gaps 36;

Qy 1013 RRRSRIMEETSGMEE-----EEEBESIAAVPQRRGRRD----- 1046
Db 9 RRKEEKLQEQKMKQEKLEDEQRNSAAVPGEEEREDPDTSTENKNIQKOLDPDVW 68
Qy 1047 -----GEVDATA--SSIPELERQI-----EKLKROQLFFRKL 1077
Db 69 TEDEDDFGSHKRSRRGVQTAVRQCIKQEBMYCIKQEPLSADAEALRQEQQKEL 128
Qy 1078 LHSSQMLRAVS----LGODRVRRYVWLPYLAGIFVEG----TEGNLVPEVIKKTDS 1128
Db 129 LDKIQSAIACNTIFPLGRDLRYRYWIFPSIPGFIEDYSGLTEDMLLP----- 178
Qy 1129 LKVAHAASLNPALFMSKVELAGSNNTASSPARASRPLKTKPGFMQPRHFKSPVRQDSE 1188
Db 179 -----RP----- 180
Qy 1189 QPQAQLQPEAQLHVPAPQPOLQLOLQSHKGFLEQEGSPLSLGQSHDLSOSAFLSWLSQ 1248
Db 181 -----SSFHNAQPRDP-QVSIKTESFLSESTS--SLDQGPFD----- 216
Qy 1249 TQSHSSLLSSSVLTPD-----SSPGKLDPAPOPEPEPEDEAESSPDLOAFWNISAQM 1303
Db 217 ---DSVLLPKPVHKPNRWCIFYSSCAQLDLDALNSRGHRESALKETLLQE-KSRICAO 272
Qy 1304 PCNAAPTPLAVSDDQTPSPQQIASSKPM-NRPSAANPCSPVQFSSTPLAGLAPKRAG 1362
Db 273 -----AHFSEKHFSDKQADSKPVSSRGSSGACDISQMSAE----- 311
Qy 1363 DPGEMPOSTGLGQPKRGRPPSKFFQMEQORYLTQTAQVPPMEMSGWMIIDPEMLD 1422
Db 312 -----RQELRL----- 318
Qy 1423 AMLKALHPRGIREKALHKLHKLHKLHKLHKLHKLHKLHKLHKLHKLHKLHKLHKLHKLH 1471
Db 319 -----RDFLLDI-----EBRIYOGTIGALIKVTDQVWRSALE 350
Qy 1472 EG-----IMSWSPKE---KTYETDLAVLQWVEELBQ---RVIMSD--LQIRGWTCPSPDST 1519
Db 351 NGRVELLSSESKENGVIKTVNED-----VEEMEMEQARVIVRDLRLGIKTET-PSITST 403
Qy 1520 REDLAYCEHLSDSQEDITWRGPGREGAPQKTNPLDLAVMLAALAEONVRRYLRREPL 1579
Db 404 -----SASTPQ-SVSNVVHYLALFQIEQGIERFLKAPL 438
Qy 1580 WPTHEVVLKALLSTPNGAPEGTTTETISYEITPRIRIWRQTLQRCRSHAAHVCLCLGHLER 1639
Db 439 -----DGNDSGRSYKTV--LDRWRRESLLSSASLSQVFLHLSTLDR 476
Qy 1640 SIAWEKSVNKVTCVLCVRKGDNDFFLLCGDGRGCHYIYCHRPKMEAVPEGDFWCTVCVLAQ 1699
Db 477 SVMKSKSILNARCKICRKGDAENWLVCDGCRGHHTYCVRPKLVKAVDGDGDFWCFCEPRK 536
Qy 1700 Q-----VEGEFTQK----- 1708
Db 537 QRSRLSSRRORPSLESDEEMEGMEDDDDEVDDEEGQSEEEYEVEQDESDSDDEAL 596
Qy 1709 -----PGPPKRGOKRKSQVSLNFSF----- 1728
Db 597 SPPKRGAPQVRLPIKTKGRFGFSPSRSQRPDGRYPSRSQOSTPKNTAKSASKNLKTR 656
Qy 1729 ----GDGRRRRVLKLG-RESPAAGPRYSEERLSPSKRRR----- 1762
Db 657 SAPPTETSLVGRSSTRHSFSAIQDVVVELLSPHSKRGRKGADHTPEHSPSFTNFRVS 716
Qy 1763 -----LSMRNHS-----DLTFCEI 1777
Db 717 TSRSRQLIPLNTAESLSLQHSERKRGRKRQSTESSVPVPLNRRSSRGQGVHLSAPEQ 776
```


D	b		218	YCFKPKMDNTPDGDWYCYECVKNATNKERKICVCGGHRPSPVGKMIYCDLCPRAYHACYI	277
Q	y		1702	-----EGEFTQKPGFPKRQKRKSGYSLNF-----SEGDGRRR-----	1734
D	b		278	PLLKVPGRKWYCHGCISRAPPKKRSAGGTSGSSSKSRDRDRDRESGSAKRRSDNSKTP	337
Q	y		1735	-----RVLLKGRES-----PAA--GPRTS-----	1751
D	b		338	AMEHMQQOQMPLAGGDSHHHHHQOPSLNSGHDESMNSLPAAPLSPAHSVVSATNYDDQH	397
Q	y		1752	-----	1751
D	b		398	HANNSVDGSRFHAHLIPPSNNNGTAALLEDPVGGANVMGVYPVPTVAAGNFSAGLINQ	457
Q	y		1752	-----	1751
D	b		458	APVQPAMPFANVAVMSPRVPTPTRTPTTPTAPTPTPPPTPLLMOASPTATAHVNAVC	517
Q	y		1752	-----	1769
D	b		518	QSPPOQQAQLMTMPPSPAIGVTATTNQMGPPPINIHAIOEAKEKLQEKKEKHATKKLM	577
Q	y		1770	SDLTFCEIILMEMESHDAANPFEPVNPRLVGYRRIIKNPMDFSTWRILLRGGYTSSE	1829
D	b		578	KELAVCKTLGEMELHEDSWFFLLPVNTKFQPTYRKIKITKPMJLSTIKKKLODLSYKTRE	637
Q	y		1830	EFAADALLVFDCNOTFNEDDSEVGKAGHIMRRPPESRWEE	1869
D	b		638	DFCVDVQRIFDNCMEFNEDDSPVGKAGHGMRKFFESRWGE	677
RESULT 30					
Q	8C8D1			PRELIMINARY;	
I	D	Q8C8D1		PRT;	718 AA.
A	C	Q8C8D1;			
D	T	01-MAR-2003	(TrEMBLrel. 23, Created)		
D	T	01-MAR-2003	(TrEMBLrel. 23, Last sequence update)		
D	T	01-MAR-2004	(TrEMBLrel. 26, Last annotation update)		
D	E	Mus musculus	10 days neonate cerebellum cDNA, RIKEN full-length		
D	DE	enriched library, clone:B33006OC03	product:hypothetical Bromodomain/		
D	DE	PHD-finger containing protein, full insert sequence.			
G	N	Name=BC065123;			
O	S	Mus musculus (Mouse).			
O	C	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
O	C	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
X	N	NCBI_TaxID=10090;			
R	N	[1]			
R	P	SEQUENCE FROM N.A.			
R	C	STRAIN=C57BL/6J; TISSUE=Cerebellum;			
R	C	MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;			
R	X	Carninci P., Hayashizaki Y.,			
R	A	"High-efficiency full-length cDNA cloning.";			
R	T	Meth. Enzymol. 303:19-44(1999).			
R	L	[2]			
R	P	SEQUENCE FROM N.A.			
R	C	STRAIN=C57BL/6J; TISSUE=Cerebellum;			
R	C	MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;			
R	X	RIKEN FANTOM Consortium;			
R	A	"Functional annotation of a full-length mouse cDNA collection.";			
R	L	Nature 409:685-690(2001).			
R	L	[3]			
R	P	SEQUENCE FROM N.A.			
R	C	STRAIN=C57BL/6J; TISSUE=Cerebellum;			
R	C	The FANTOM Consortium,			
R	A	the RIKEN Genome Exploration Research Group Phase I & II Team;			
R	T	"Analysis of the mouse transcriptome based on functional annotation of			
R	L	60,770 full-length cDNAs.";			
R	T	Nature 420:563-573(2002).			
R	L	[4]			
R	N	SEQUENCE FROM N.A.			
R	P	STRAIN=C57BL/6J; TISSUE=Cerebellum;			
R	C	MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;			
R	X	Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,			

Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
 "Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new genes.";
 Genome Res. 10:1617-1630(2000).
 [5]
 SEQUENCE FROM N.A.
 RP STRAN=C57BU/6J; TISSUE=Cerebellum;
 RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
 RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
 RA Konno H., Akiyama J., Nishi K., Kikunai T., Tashiro H., Itoh M.,
 RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
 RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
 RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watanishi M.,
 RA Onoda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J.,
 RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
 "RIKEN integrated sequence analysis (RISA) system-384-format
 sequencing pipeline with 384 multicapillary sequencer";
 Genome Res. 10:1757-1771(2000).
 [6]
 SEQUENCE FROM N.A.
 RP STRAN=C57BU/6J; TISSUE=Cerebellum;
 RC Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
 RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
 RA Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,
 RA Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
 RA Katoh H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,
 RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
 RA Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,
 RA Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
 RA Sasaki D., Shibata K., Shingawa A., Shiraki T., Sogabe Y., Tagami M.,
 RA Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,
 RA Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;
 Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
 RL EMBL: AK047427; BAC33055.1; -
 DR HSP; Q9UIG0; Ipf2.
 DR MGD; MGI:3039627; BC065123.
 DR GO; GO:0001151; C:ubiquitin ligase complex; IEA.
 DR GO; GO:0003677; F:DNA binding; IEA.
 DR GO; GO:0004842; F:ubiquitin-protein ligase activity; IEA.
 DR GO; GO:0008270; F:zinc ion binding; IEA.
 DR GO; GO:0016567; P:protein ubiquitination; IEA.
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
 DR InterPro; IPR001487; Bromodomain.
 DR InterPro; IPR011011; FYVE PHD Znf.
 DR InterPro; IPR001965; Znf_PHD.
 DR InterPro; IPR001841; Znf-ring.
 DR Pfam; PF00439; Bromodomain; 1.
 DR Pfam; PF00628; PHD; 1.
 DR PRINTS; PR00503; BROMODOMAIN.
 DR SMART; SM00297; BROMO; 1.
 DR SMART; SM00249; PHD; 1.
 DR SMART; SM00184; RING; 1.
 DR PROSITE; PS00633; BROMODOMAIN_1; 1.
 DR PROSITE; PS0014; BROMODOMAIN_2; 1.
 DR PROSITE; PS01359; ZF_PHD_1; 1.
 DR PROSITE; PS00016; ZF_PHD_2; 1.
 KW Hypothetical protein
 SQ SEQUENCE 718 AA; 81471 MW; B508483BEB24FE02 CRC64;
 Query Match 4.3%; Score 425.5; DB 2; Length 718;
 Best Local Similarity 22.4%; Pred. No. 1e-09;
 Matches 167; Conservative 86; Mismatches 179; Indels 315; Gaps 24;
 Qy 1368 PQSPTGLGQPKRGRPPSKFFKQMEQRYLTOLTAQ-----PVPPECMCGWWI 1415
 Db 6 PSSFHNAQP-----RDFQVSIK-TESFSJSESTSLDQGFDDSVLLPKPVHKPNRCWFY 60
 Qy 1416 PDPEMLDAMLKALHPGRIKALHK-----HLNKH-----1445
 Db 61 SSCAQLDQLDALNSGRHSALKEKTYLLOEKSPICLAHLFSEKHFSDKPOADSKPVS 120
 Qy 1446 -----RPFLOEVCLRPSADPIFE-----PQQL--PAF 1470


```
Db 121 SRSSGACDISQMSAEROLELRDRFLLDI-----EDRIYOGTIGAIVTDQRQWRGAL 175
QY 1471 QEG-----IMSWSPKE---KTYETDLAVLQWVELEQ---RVMSD--LQIRGWTCPSPDS 1518
Db 176 ENGRYELLSEESKGVITVNE-----VEEMEMEQAQVIVRDRLLGIKIKTET-PSITIS 228
QY 1519 TREDLAYCEHLSQSDQEDITWRGPGREGGLAPQKRTTNPLDLAVMLAALEONVRRYLRBP 1578
Db 229 T-----SASTPQ-SVSNVVHYLALFOIEQGIERRFLKAP 263
QY 1579 LWPTEHVLEKALLSTPGAPGTTETLSYEITPRIRIWRQTLQCRSAAHVCLCLGLE 1638
Db 264 L-----DGNDSGRSYKTV--LDRWRESLLSASLSQVFLHLSTLD 301
QY 1639 RSIANEKSVKVTCLVCRKGNDFLLLCDGDRGCHYCHRPKMEAVPEGDFWCTVCIA 1698
Db 302 RSVMSKSLNARCKICRKGDAENWVLCGDRGHYTCVVRPKLKAVIDGDMFCECRP 361
QY 1699 QO-----VEGEFTQK-----1708
Db 362 QORSRLSSRQPSLESDEEMEGMEDDDDDDEVDDEEGQSEEBEYVEQDEEDSDDEA 421
QY 1709 -----PGPPKRGOKRKGVSINFSB-----1728
Db 422 LSPPKRGPRQVPLPTKTKRGFCSPFSRSQRPGRYPFSRSQOSTPKNTAKGASKNLRKT 481
QY 1729 -----GDGRRRVLKGG-RESPAAGPRYSERLSPSKRR-----1762
Db 482 RSAPTETRLVGRSSTRHSFSLQDVVELLSPHSKRGRKGADHTPEHSPSTFNFRV 541
QY 1763 -----LSMRNHS-----DLTCE 1776
Db 542 STSRSSROLIPLNTAESLSQHSSESKRRKRKRQSTESSPVPLNRRSSGRQGVHLSAFE 601
QY 1777 IILMESHDAWPPKXPVNPRLVSGYRRRIKNPDMFSTMRERLLRGVTSSEFAAAL 1836
Db 602 QLVVELRHDSWPFLKLVSKIQVPDYDIIKKPIALNIIRKVNKCEYKLASEFIDIE 661
QY 1837 LVFDMCOTFNEDDSVGVKAGHIMRFF 1863
Db 662 LMFSNCFEYNPRNTSEAKAGTLOAFF 688
RESULT 31
ID Q8COK4 PRELIMINARY; PRT; 861 AA.
AC Q8COK4;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Mus musculus adult male thymus cDNA, RIKEN full-length enriched
DE library, clone:5830435C13 product:BROMODOMAIN ADJACENT TO ZINC FINGER
DE DOMAIN 2B (Fragment).
GN Name=5830435C13Rik;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Thymus;
RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning.";
RL Meth. Enzymol. 303:19-44(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Thymus;
RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RA RIKEN FANTOM Consortium;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
RN [3]
```

```
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Thymus;
RA The FANTOM Consortium;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Thymus;
RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RA Carninci P., Shibata Y., Hayateu N., Sugahara Y., Shibata K., Itoh M.,
RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT "Normalization and subtraction of cap-trapper-selected cDNAs to
RT prepare full-length cDNA libraries for rapid discovery of new genes.";
RL Genome Res. 10:1617-1630(2000).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Thymus;
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RA Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M.,
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA Fujiwara S., Inoue K., Togawa Y., Izawa M., Ohara E., Watanabe M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
RT sequencing pipeline with 384 multicapillary sequencer.";
RL Genome Res. 10:1757-1771(2000).
RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Thymus;
RA Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
RA Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,
RA Hori F., Imokani K., Ishii Y., Itoh M., Kagawa I., Katsukawa T.,
RA Katoh H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,
RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
RA Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,
RA Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
RA Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,
RA Tagawa A., Takahashi F., Takaku-Akashira S., Takeda Y., Tanaka T.,
RA Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK030867; BAC27164.1;
DR MGD; MGI:2442782; 5830435C13Rik.
DR GO; GO:0003677; F:DNA binding; IEA.
DR InterPro; IPR001739; Methyl-CpG_bind.
DR Pfam; PF01429; MBD; 1.
FT NON TER 861
SQ SEQUENCE 861 AA; 92757 MW; 10AE56161FDA4626 CRC64;
```

```
Query Match 4.1%; Score 407.5; DB 2; Length 861;
Best Local Similarity 21.1%; Pred. No. 7e-09;
Matches 184; Conservative 123; Mismatches 299; Indels 267; Gaps 27;
QY 16 GLPP--APASGLKPSPPSGGLYTNGSPMFPQOGKSLNGDVVYGLSTVSTHTTSGIL 73
Db 89 GIPPLFAPPQN---HDSSTHRTSGKSRNGPEKG--ING--SVNGTSAASVLGVNASV 141
QY 74 NSAPHSSSTLHLHPVSAYDCLWNYSPQPSANP-----106
Db 142 VATPASSMGQNGQSTSGGGTLKCHQEQNKQSPVDARADRIKDKPKRKAMESSNSDSD 201
QY 107 -----GSNLKDPPLLSQ-----FSGQVPLNGILGG 132
Db 202 SGTSSDTSSEGISSSDSDSDDLEEBEEDQSVSEEDDDSDSETEAQHKNNQVLLHGI--- 258
QY 133 SRQSSPSHNTNLRAGSKFWANGTHSPM-----GLNFDQSRLYDSFPDQNEEVCVS 184
Db 259 ----SDP--KTDGQKATEKAQERRTHQPLPPVSESQTHPPFQSQQKQPVLSQQLFFIQ 312
```



```
Query Match      4.0%; Score 396.5; DB 1; Length 627;
Best Local Similarity 23.0%; Pred. No. 1.3e-08;
Matches 145; Conservative 73; Mismatches 160; Indels 253; Gaps 20;

Qy 1451 EVCLRPSADPT-----FPPQLPAPQEGIMSWPKETETDYLAVLQW 1493
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3 KICQLDSFFVNRFRNIADNPTPLDFKPPFRGLKLNASGSPVS--AEKLELRCLDL-- 57

Qy 1494 VEELEQRVI-----MSDLQRCWTCPSPDSTREDL-----AYCEHLSQEQ 1534
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 58 ILDTEDRIYQGLGAVKVPDRQL--WRSALENNNTLELNDKAKENGSSKPINHEMEEMI 115

Qy 1535 DITWRGPGR-----EGLAPQKTTNPLDLAVMRLA---ALEQNVKRRYLREPLWP 1581
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 116 DIKOSAKDRLLGLKNDTPSATSTSTNPQPVNNAVRYLARALHOIQGVKRYKAPL-- 173

Qy 1582 THEVLEKALLSTPNGAPEGTTTBISEYITPRIMRQTLQRC-----RSAHVCLCL 1634
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 174 -----GDASEG-----RVQRTVLDRCGESLSSGSPQIFLHL 207

Qy 1635 GHLERSIAWKSNNKVTCLVCRKGDNDFFLLCDGCRGCHYIYHRPMEAVPEGDFCT 1694
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 208 STLDRLSILWRSILNARCKVCRKKGDSVMVLCDCDRGHHIYCVRPKLYVPEGDFCP 267

Qy 1695 VCLAQQ-----VEGEFTOKPG 1710
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 268 ECHPKQKSHRLPSRHRYSMDSEEEELDQKEEEEBEERQESSENGQEDMSEES 327

Qy 1711 FPKR-----GOKRKS----- 1720
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 328 PPKEGRAKVLPLKWRGKATGLPKPKTKGQSTPKNTQPAPEGROGKTRSAPSLEP 387

Qy 1721 -----GVSIMFS-----EGGRRRRVLLKGR-SPAAGP----- 1748
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 388 TSRLSASDSPAHGVSPNSLNVVTVNGRGGKGRGRGRLOKSADNTPASSPAFR 447

Qy 1749 -----RYSEERLSPS-----KRR-----RLSNRNHS--DL 1772
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 448 PFLSDSNEPTPPGRKPSQSLLPSQAPKGGKRLSGDISLSLEQGNRRSGRHHGVHEL 507

Qy 1773 TFCBIILMEMESHDAAPFPXEPVNPRLVSGYRRIIKNPMDFSTWRRLRLLRGYTSSEFA 1832
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 508 SACEQLVELVRHDSFPFRLVSKNQVDPYFDVIQRIALNLRKVNRCYKCASEFI 567

Qy 1833 ADALLVFNCOTFNEDSEVCKAGHIMRRFF 1863
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 568 DDVHLMNSCFEYNHNSNEAKGIRLQSPFF 598

RESULT 34
Q7Z7D6 PRELIMINARY; PRT; 2764 AA.
AC Q7Z7D6;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Bromodomain PHD finger transcription factor.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Barak O., Lazzaro M., Lane W., Speicher D., Picketts D.,
RA Shiekhattar R.;
RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY282495; AAP22284.1; -.
DR HSSP; Q92830; 1F68.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR001487; Bromodomain.
DR InterPro; IPR000345; CytC_heme_BS.
```

```
DR InterPro; IPR004022; DDT.
DR InterPro; IPR006209; EGF like.
DR InterPro; IPR011011; FYVE_PHD_Znf.
DR InterPro; IPR001965; Znf_PHD.
DR Pfam; PF00439; Bromodomain; 1.
DR Pfam; PF02791; DDT; 1.
DR Pfam; PF00628; PHD; 2.
DR SMART; SM00297; BROMO; 1.
DR SMART; SM00571; DDT; 1.
DR SMART; SM00249; PHD; 2.
DR PROSITE; PS00633; BROMODOMAIN 1; 1.
DR PROSITE; PS00014; BROMODOMAIN 2; 1.
DR PROSITE; PS00190; CYTOCHROME_C; UNKNOWN 1.
DR PROSITE; PS00827; DDT; 1.
DR PROSITE; PS01186; EGF_2; UNKNOWN 1.
DR PROSITE; PS01359; ZF_PHD 1; 2.
DR PROSITE; PS00016; ZF_PHD 2; 2.
SQ SEQUENCE 2764 AA; 308079 MW; 9785D7097C2099F9 CRC64;

Query Match      3.9%; Score 383; DB 2; Length 2764;
Best Local Similarity 18.0%; Pred. No. 3e-07;
Matches 352; Conservative 271; Mismatches 743; Indels 590; Gaps 72;

Qy 79 SSSTSHLHPSVAYDCLMNVSYQPSANPGSNLKDPLLSPGQYPLNGILGSRPSS 138
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1234 SSSKSALH-----SSVPKS---TNDRDATPLSR---AMDPEGKLGCDSESN 1274

Qy 139 PSHTNLRAGSQFWANGTHSPMGLNFDSELVSDPDQNFEEVCSGIHDEAAEKWTS 198
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1275 TLENSDVTSIQ-----DSSSEDMIVQNSNESISEQFRTRE-QDV-----EVLEPKCE 1322

Qy 199 VVAENGVLCSLELEKQPELKMCGYNGSVPSVESLHQVSVLVPDPTVSCLDPSHLP 258
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1323 LVGESTG-----NCDRLP-VKGTENGKKPQOKKLEE-----RPVAKCSQ----- 1365

Qy 259 DQLEDTPILSEDSLEFPNSLAPEFVSGLYGIDDTLMGAEDKLP-----LXDS 307
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1366 -----IKLNTTDDKKNENRESEKKGQR--TSTFQINGKDNKPKIYLKGECLKSEIS 1416

Qy 308 PVISALDCPSLNATAPSLADDSQTSIFASPTSPVLGESVLQD-----NSFD----L 359
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1417 RVVSGNVPEPVKNIN--KIIPENDIKSLTVKESAIRPFINDGVIMEDFNERNSETKSHL 1474

Qy 360 NNGSDAE--QEWME---TQSSDFPPSLTOPA--PDQSTIQQLHP---ATPAVSPTTS 407
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1475 LSSSDAEGNVRDSLETLPSTKESDSTQTTTPSACPSNVQVEDMEIETSEVKVTSS 1534

Qy 408 PAVSLVVSPPAASPEIS-----PEVCPAAGTVV--SPAVFSVVS 443
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1535 PITSEESNLNDIFENGLPINKNENVNGESKRKTVITEVTTMTSTVATESKTVIKVEK 1594

Qy 444 PASSAVLPAVSLEVLPTASTVSPKASPTSPAAAFPTASPAKDVSSFLETTADVEITG 503
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1595 GDRQTVVSSSTENAKSTVTTTTTIVTKLSTPSTGGSVDIISVKEQSKTVTTVTDSLTT 1654

Qy 504 EGLTASGSDVMRRRIATPEEVRPLQHGWRREVRIRKKGSHRWQGETWYWGPCGKRMKQF 563
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1655 TGGTLVTSMTV--SKEYSTRDKVKL---MKFSRPKKTSGT-----AL 1692

Qy 564 PEVIKYLNRNLVHSV-----RRE--HFSP-----SPRMPVGD 593
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1693 PSYRKFTVKTKKSI FVLPNDDLKCLARKGIRVYFYNNAKPALDIWPVPSRPRTFGI 1752

Qy 594 FFEER-DTPEGLQ-----WVLSAEHIPSRIQAITKGRGRPRNTEKAKTKEVPKVK 643
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1753 TWRYRLQTVKSLAGVSLMLRLLLWASLRWDDMAAKVP---PGGSTRTETSET----- 1801

Qy 644 RGRGRPEPKVITELLNKTDRP-----LKKLEAQETLNEEDKAKIAKSKKMRQKVQ 695
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1802 -----EITTTETILKRRDVGPIRVEYCIKIKIICPIGVETPKETPTPORKGLRSSAL 1854

Qy 696 RGECLTTIQGQARNKRKQETKSLKHKEAKKSKAEKKGKTKQBLKEKVKREKKEKVKM 755
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```


DR SMART; SMO0249; PHD; 2.
DR SMART; SMO0184; RING; 2.
DR PROSITE; PS00014; BROMODOMAIN 2; 1.
DR PROSITE; PS00119; ZF_BOX; 2.
DR PROSITE; PS01359; ZF_PHD 1; 1.
DR PROSITE; PS00016; ZF_PHD 2; 1.
DR PROSITE; PS00518; ZF_RING 1; 1.
DR PROSITE; PS00089; ZF_RING 2; 1.
KW Metal-binding; Zinc; Zinc-finger.
FT NON TER 1
SQ SEQUENCE 1071 AA; 116872 MW; C1DD015CE156A4C5 CRC64;

Query Match 3.6%; Score 353; DB 2; Length 1071;
Best Local Similarity 19.4%; Pred No. 1.6e-06;
Matches 241; Conservative 149; Mismatches 400; Indels 450; Gaps 50;

QY 807 PLPDFSRVPLGTLPSGAFSDCLT-----IVFSLFSFGVLGDFPAKQVPSL 852
DB 67 PAPGSSGPPLPASPASLLDTCVACQSLQSRREAEKLLPCLHSFCLRLCLPEPER----- 121
QY 853 GVLEQGLLCOGDSLGEVDLRLVLLKAAHLDPGPPSYCQSLKILG-----EKVSEIPLTRD 908
DB 122 ---QLSVPIPGSGNDVQV-----GVIRCPVCRCQCIDLVNDYFVKDTSEAPSSD 172
QY 909 NVSE-----ILRCFLMAY-----G 922
DB 173 EKSEQVCTCEDNASAVGFCVECEWLCCKIEAHQVRKTKOHLIRKEDVSESVTSG 232
QY 923 VXPALCDRLRTQFPQAQPPQQAALVAPFVHELNGSTLIINEIDK-----TLESM--- 972
DB 233 QRPVFC-----PVHKQQLKLCFCETCDRLTCRCQLLEHKEHR 270
QY 973 -----SSYRNKWIVEGRRLRLKTVLAKRT-----GRSEVEMGRPECLGRRSRIME-- 1021
DB 271 YQFLEEAQKQGAIENTLAKL-----LEKQNVHFAATQV-----NRIKEVN 314
QY 1022 ETSQMEEEEEEESTAAVPGRRGREGDGVDTASSIPELERQIEKLSKQLFRFKLLHSS 1081
DB 315 ETNKRVEQEIKAIVFTL-----INEINKGKS-----LLOOLENVTKER---QMKLQQQ 361
QY 1082 QMLRAVSLGQDRYRRRYWVLPYLAGIFVEGTGNLVPEE-----VIKKTDSLKVA 1132
DB 362 NDITGLS-----RQVKHVMFTWATASGSTALLYSKRLITPQLRHILKARCDPVPA 415
QY 1133 AHA---SINPALFPMKMLAGSNTTASSAPARSRPLKTPGF----- 1172
DB 416 NGAIRPHCDPTFWAKNVNGLNVLIESKPA-----PGYTPNVVVGQVPPGTNHIS 465
QY 1173 -----MQPRHFSPVRGQDSEQPAQLQPEAQLHVPAPQPPQLQLQSHKGFLE 1222
DB 466 KTPGQINLAURLQHMQQVVAQKHQQLQ-----QMLQOPPAPIPITTTATTQOH---P 516
QY 1223 QEGSPLSLGQ-----SQHDLQSAPFLSMLSTQGHSSLLSSVLTPDSSPGKLD 1271
DB 517 ROAAPQLQOQPPRLISVQITWQRCNMCGAF-----QAHQMLAQNAARIPGIP--RH 567
QY 1272 PAPSQPPEEPEDAEESPDLOAFWNISAMPNCN-AAPTPPLVASEDQPTSPSQQLA-- 1328
DB 568 SAPQYSMMQHLQHQSNPGHAGFPVVSANHPNINPTSPATTMANANRGPTSPSVTAIE 627
QY 1329 -----SSKPMNRPSAANPCSPVQF-----SSTPLAGLAPKRAG-----DPGEMP 1368
DB 628 LIPSVTPENLPSPUD-IPPIQLSDAGSSSLDNLSRYISGSHLPPOPTSTMNPSGPSPA 686
QY 1369 QSP--TGIGLQPKRRGRPPSPKFFKQMEQRYLTQLTAPVPPVPMCMGWWIIPDPEMLDAMLK 1426
DB 687 LSPGSSGLNSHTVPRPS-----TSSTGS----- 711
QY 1427 ALHPRG-----IREKALHKLHKNHRDFLQEVCLRPSADPIFPRQLPAPQEGIMGNWP 1479
DB 712 ----RGSCGSGSRGTAESKSAHSF-----KSDQV-KVKQEPGPTGEEICFSFG 751
QY 1480 KEKTYETDLAVLOWVEELQVRVMSDLQIRGWTCPSPDSTREDLAYCHELSDSQEDITWR 1539

DB 752 AVQEKTE-----DGRRSACMLSSPSS----- 774
QY 1540 GPGREGLAPORKTNTPLDLAVMLRLAALQNVKRYLREPLWPT--HEVWLEKALLSTPNG 1597
DB 775 -----LTPPLSTNLHLESELDLTGLGNHVKTE-----PTDISCSCKQSGLSNLVNG 821
QY 1598 APESGTTTSEIYEITPRIRIWRQTLQRCRSHAAHVCCLGHLERSIAWEKSVNKKVTCCLVCRK 1657
DB 822 K-----SPIRNLHRSARIGGCGNS--KDDDPNEDWCAVCQN 856
QY 1658 GDNDEFLLCDGCDRGCHIYCHRPQMEAVPEGDFTCVCLAAQVVEGEFTQKPGPPKRGQK 1717
DB 857 GGD---LLCCCKCPKPVHLTCHVPTLLSPFGDWICTFC-----RDIGKPE--- 899
QY 1718 RKSGLSYNFGSGDGRRRVLLKGRSPAAGPRVSEERLSPSKRRRLSRNHHSLDTFCEI 1777
DB 900 -----VEYDCDNMQHSHKKGK-----TAQGLSPVDQRK-----CER 929
QY 1778 ILMEMESHDAWPEXEPVNPRLVSGVYRIIKNPMDFSTMRERLLR---GGYTSSEFAAD 1834
DB 930 LLLLYLYCHEUSIEFQEPV-PVSIPTYIKLIKPMDLSTVKKLQKKHISOHYQIIPDDFVAD 988
QY 1835 ALLVFDNCQTFNEDDSEVGKAGHIMRRFFESRWEEFYQK 1874
DB 989 VRLIFKNCFERNEGDSSEVAKAGAVLYFEDKLSIYSDR 1028

RESULT 38
Q924W6
ID Q924W6 PRELIMINARY; PRT; 1209 AA.
AC Q924W6;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE KIAA0298 protein.
GN Name=Kiaa0298;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
[1]
RN SEQUENCE FROM N.A.
RP MEDLINE=21418998; PubMed=11528127;
RA Amid C., Bahr A., Mujica A., Sampson N., Bikar S.E., Winterpacht A.,
Zabel B., Hankeln T., Schmidt E.R.;
RT "Comparative genomic sequencing reveals a strikingly similar
architecture of a conserved syntenic region on human chromosome
11p15.3 (including gene STS) and mouse chromosome 7.";
RL Cytogenet. Cell Genet. 93:284-290(2001).
DR EMBL; AJ307670; CAC38114.1; -.
DR HSBP; Q14839; lMW3.
DR GO; GO:0005622; C:intracellular; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR Pfam; PF00439; Bromodomain; 1.
DR Pfam; PF00628; PHD; 1.
DR Pfam; PF00643; zf-B box; 2.
DR PRINTS; PR00503; BROMODOMAIN.
DR SMART; SM00502; BBC; 1.
DR SMART; SM00336; BBOX; 2.
DR SMART; SM00297; BBOX; 1.
DR SMART; SM00249; PHD; 2.
DR PROSITE; PS00014; BROMODOMAIN 2; 1.
DR PROSITE; PS00237; G PROTEIN RECEPTOR 1; UNKNOWN 1.
DR PROSITE; PS01119; ZF_BOX; 2.
DR PROSITE; PS01359; ZF_PHD 1; 1.
DR PROSITE; PS00016; ZF_PHD 2; 1.
SQ SEQUENCE 1209 AA; 133324 MW; A09AE83390F46D6C CRC64;

Query Match 3.6%; Score 350.5; DB 2; Length 1209;
Best Local Similarity 20.5%; Pred. No. 2.4e-06;

	Matches	241;	Conservative	130;	Mismatches	377;	Indels	425;	Gaps	50;
Qy	966	DKTLESMSSYRKNKIWEGRLRRLLKTVLAKRTGRSEVMGRPCECLARRSRRIEMSTSG	1025							
Db	133	DRIFEVXHOHRK---VENQIKMAKWMLNEL-----NKQANGLEELEG	173							
Qy	1026	MEEEEBEESIANVPGRGRRRDGEVDATASSIPELERQIE-----KLSKROLFF	1073							
Db	174	ITNE-----RKRLKEOQLQSIMVLNRQFHVQNFINWAVCSKSSVPFLFS	218							
Qy	1074	RKKLLHSSQMLRAVSLGD-----RYRRYYVV--LPYLAGIFVEGTE-----	1113							
Db	219	KELIVFOWORLETRCWTDPSPWSIRIPTWEPNFTWKQLASGLCITTEGGOLTRADAAA	278							
Qy	1114	--GNLYPE-----EVIKKETDSLKVAHAASLNPAFLSMKMELAGSNNTTASS-----	1157							
Db	279	SYGSLOGOPSFYQSHQAQPMACQEALSHPSHKFKQSPALCSSVCSCSHCPSPVPSLKGVPP	338							
Qy	1158	----PARARSPLKT-----PAPRMQPRHFPSVRVGDD--	1186							
Db	339	PSIHAFHSFRQPFSEMVPHQLGSLQCSTLLPREKELACSPHPKLMQWLPEQPQPAEQEST	398							
Qy	1187	SEQQAOLAQPEAQLHP-----AQP--QPLOLOLOSHK-----GFLEQ-----	1223							
Db	399	SOREGPQLVSQPCIVPDQVQGAHAQPTIQTPSIQVLGHOKLKLSHFQQQPOQOPP	458							
Qy	1224	-----EGSPLSLGQSQH-----DLSQSFAF-----LSWLSQ	1248							
Db	459	PPPPPPPQHAPPPLPPSQHLASSOHESPAPACSQNVDMHHKFLEEMQKDLELLQL	518							
Qy	1249	TQSHSSLSSSVLPDSDSPGKL-----DPAPSQPPEEPDEPAESSPDLOAQF	1295							
Db	519	AQPSLQLSQT-----KSPHQLQTIIVGQINYIVRQAPVQSOSQEBTLQVTTEP----	568							
Qy	1296	WFNISAQMCWAAPTTPPLAYSEDQPTSPPOOLASSKPKWRPSA-----AN	1340							
Db	569	-----PAPEGKPALPVDKNTAAPLPOTSGBETHSPVDPVDTGSHSSPNVVRKHAT	620							
Qy	1341	PCSVPQFSSTPLAGLAPKR--RAGDP-----GEMPOSPTGLGQPKRGRRPSPKFFKOM--	1391							
Db	621	SVSIMGSENTVEMELSSTRLATETPIQIHRVSSSTAAPTHT--IPSLLSGGPPTVSSLSMV	679							
Qy	1392	EQRILTOLTA--OPVP-----PMSQGMWWMP	1416							
Db	680	SNHAMPSLTASHLQPVNLVRGTFTOSTNLRGDSQAITGLASNHSQAQPSLMSGHTQAA	739							
Qy	1417	DPEMLDAMLKAHP-----RGIRE-----KALKH-----	1440							
Db	740	-PSLATCFLOQMPPVSDVHVPRSVSSPGSPAAESIGTRDGAESSLGNALCKVELPEPI	798							
Qy	1441	HLNKHDFLOEVLURPSA-DPIPEPROLPAPFOEGIMSWSPKETYETDLAVLQWVELEQ	1499							
Db	799	NLSVKKFP LAPINTSTALOQYRNPKYENFEQGALELDTKENS-----DI	844							
Qy	1500	RVIMSDLQIRGWTCPSDSTREDIAYCEHLSDSOEDITWRGPGREGIAPORKTTNPILDLA	1559							
Db	845	RAISSEPFI-----PVRRLERLKIC-AASSGEMPVFK-----LKPQKSQODGNFLL	899							
Qy	1560	VMLLAALEQNVKRRYLEBPLWPHEVLEKALLSTPNGABEG-----TTTTISVIETPRIR	1615							
Db	890	VIECGTESSMSIK-----VSQNSLPDASQGPGIGLRKVTVTSLT-----	929							
Qy	1616	IWROTLQCRSAAHVCLIGH--LERSIAEKSV-----NKVTVLCVCRKGDNDFFLLCD	1668							
Db	930	--GOQPEVETSE-----EHRLLIPRAPGAKNTPAPIENEDFCVACINGGE---LLCCD	979							
Qy	1669	GCDRGCHYCHRPKMCAVPEGDMFTCVCLA-QQVEGEF---TQKPGPPKGQRKSKGYSL	1724							
Db	980	RCPKVTHLSCHVPALLSFPGEWCTLCRSITQPEMYDCENARYGHP--GVRVPLGLSM	1037							
Qy	1725	NFSEGDRRRVLLKGRESAAGPRYSERLSPSKRRRLSRNNHSDLTTFCEILMEMES	1784							
Db	1038	-----YDQKK-----CEKLVLSLCC	1052							

```

Qy 1785 HDAWPFKPEVNPRLVSGYRRIIKNPMDFTMRERLLR---GGYTSEEFAADALLVFDN 1841
Db 1053 NLSLSPFHEFVSP-LARHYVYIIIKRPMDLSIIIRKLOKQDPAHYTTPEYVSVRLMFWN 1111
Qy 1842 CQTFNEDDSEVGKAGHIMRRFFESRWEFYGK 1874
Db 1112 CAKENYFDSVAEAGRCLEVFVFEGLKEIYPDK 1144

RESULT 39
Q9BQ04 PRELIMINARY; PRT; 1214 AA.
AC Q9BQ04;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE KIAA0298 protein.
GN Name=KIAA0298;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_taxid=9606;
[1]
SEQUENCE FROM N.A.
RP MEDLINE=21418998; PubMed=11528127;
RX Amid C., Bahr A., Mujica A., Sampson N., Bikar S.E., Winterpacht A.,
RA Zabel B., Hankeln T., Schmidt E.R.;
RT "Comparative genomic sequencing reveals a strikingly similar
RT architecture of a conserved syntenic region on human chromosome
RT 1p15.3 (including gene ST5) and mouse chromosome 7.";
RL Cytogenet. Cell Genet. 93:284-290(2001).
DR EMBL: AJ400879; CAC35389.1; -.
DR HSP; Q9UIG0; IP62.
DR GO: GO:0005622; C:intracellular; IEA.
DR GO: GO:0003677; F:DNA binding; IEA.
DR GO: GO:0008270; F:zinc ion binding; IEA.
DR GO: GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR001649; Bbox C.
DR InterPro; IPR001487; Bromodomain.
DR InterPro; IPR011011; FYVE_PHD_Znf.
DR InterPro; IPR000276; GPCR_Rhodpsn.
DR InterPro; IPR000315; Znf_PHD.
DR Pfam; PF00439; Bromodomain; 1.
DR Pfam; PF00628; PHD; 1.
DR Pfam; PF00643; zf-B_box; 2.
DR PRINTS; PR00503; BROMODOMAIN.
DR SMART; SM00502; BBC; 1.
DR SMART; SM00336; BBOX; 2.
DR SMART; SM00297; BROMO; 1.
DR SMART; SM00249; PHD; 2.
DR PROSITE; PSS0014; BROMODOMAIN 2; 1.
DR PROSITE; PS00237; G_PROTEIN_RECEP_Fl_1; UNKNOWN_1.
DR PROSITE; PSS0119; ZF_BBOX; 2.
DR PROSITE; PSS01359; ZF_PHD_1; 1.
DR PROSITE; PSS00016; ZF_PHD_2; 1.
DR PROSITE; PSS00016; ZF_PHD_2; 1.
SQ SEQUENCE 1214 AA; 134403 MW; A3973C0EA2D47C3A CRC64;

```

Query Match	3.5%;	Score 348;	DB 2;	Length 1214;
Best Local Similarity	21.2%;	Pred. No. 3e-06;		
Matches 243;	Conservative 133;	Mismatches 408;	Indels 360;	Gaps 54;
Qy	966	DKTLEMSYRNKWI	VEGRRLKTKTVLAKRTGRSEVEMGRPECLGRRSSRIMEETSG	1025
Db	133	DRIFEVKQHRK	---VENVIKWAKVNNEL-----	NKQANGLIELEG 173
Qy	1026	MEBEEBESIAA	VEGRRGRDGEVDATASSIPELERQIE-----	KLKSRQLFF 1073
Db	174	ITNE-----	---RKKLEQOLOSIMVLRQFEHVQFINWAVCSKTSVPLFLS	218
Qy	1074	RKXLLHSSQMLRAVSLGQD-----	---RYRRRYWV---LPYLAGIFVEGTE-----	1113

Db 219 KELIVFQMRLLTSCNTDPCGSPWSIRFTWEPNFWTKQLASGLCITTEGGQMSRADAPAY 278
Qy 1114 GNLVPEEVIKK-----ETDSLKVAHAASLNALFSMKWELAGSNNTTASPARSRPL 1166
Db 279 GGLGSGSPFYQSHOSPVAQQALSHPSHKFOSPAVCS--SSVCCSHCSFVPSLSLKGQVP- 335
Qy 1167 KTKPGFMQPRH-FKSP-----VRGQDSEQOQALQPEAQL-----HVPAAQOPQLOLQLQSH 1217
Db 336 ---PSPHAPHSFRQPEWVQOQLGSLQCSALLPREXELACSPHPKLLQPWLETQPP-- 390
Qy 1218 KGFLQEGSGPSLSLQCS-----QHDLSQSAPLSWLSQTS-----HSSLSS-- 1258
Db 391 ---VEQESTSRLQGLTSPQVCIVPPQDVQOGAHAQPTLTQPSIQVQFGHHQKLKLSHF 447
Qy 1259 -----SVLTDPSSFGKLDPA-P-SOPPEPEP-----DEAESSP-----DLQA 1294
Db 448 QQQPQQQLPPPPPLPHPPPPPLPPPPQHPPLPPPSQHLASSQHESPFGPACSNQMDIMH 507
Qy 1295 FWFNISAQMP-----CNAAPTPLAVSEDOPTPSPOQLASSKPMNRPSAANPCSPVQFSST 1350
Db 508 HKFELEBQKDLLELLLOAQPSLSLQ---TKSFQHLQQTIVGQINIVRQPAVPQSQSQ 564
Qy 1351 PLAGLAPKRRAGDPGEMQSP-----TGLGQPKRRGR-----PP----- 1384
Db 565 EETLQATDEPPASQSGKPALPLDKNTAAALPQASGEETPLSPVPDSTIQHSSPNVVRKH 624
Qy 1385 -----SKFFQMQORVLTQTAQVPPE-----MCSGWWIIPD--PEMLDA---MLKA 1427
Db 625 STSLSIMGFNTLEMSLSTLERPLEFQIQSVSNLTAG---APOAVPSLLSAPPKRVSS 681
Qy 1428 LHPRGIREKALHKLHGRDPLQVCLRPS-ADPFE-----P-----PRQ 1466
Db 682 L--TSVQNAQWPSLTTSH---LQTV---PSLVHSTFGSMPLNLSDSFOAMASLASDHPQA 733
Qy 1467 LPAFQEGIMSWSPKETYETDLAVLOWVEELEQRVIMSDQIR--GWTCP--SPDSTRED 1522
Db 734 GPSLMSGHTQAVPSLAT--CPLQSIPPVSDMQPETSSTGSSSSGRTSGSLCPRDGADPSLEN 791
Qy 1523 LAYCEHLSDSQEDITWGPREGIAPQKKT-----NPLDLAVMLALAEQNVKRR-- 1573
Db 792 -ALCKVKLEBPINLSVKKP---PLAPVVSTSTALQYQNPKECENFEGQALEDAKENQS 847
Qy 1574 -----YLR-----EPLW-----PTHEV 1586
Db 848 IRAPNSEHKIPYVRLERLKICAASSGEMPVKLPQKNDQDGSFLLIIECTESSSSSIK 907
Qy 1587 LEKALLTPNGAP--EG---TTTISYEITPRIRWQTQRCRSAHVCLCLGHLERSI 1641
Db 908 VSQDRLSEATQAPGLEGRKVTVTSLAQRPPEVEGTSPPEHRL-----IPRTP 955
Qy 1642 AWEKS-----VNKYTCLVCRKGDNDDEFLLLCDGDRGCHIYCHRPKMEAVPEGDFWCTVC 1696
Db 956 GAKKGPPAPTIENEDFCVCLNGGE---LLCCDRCPKVFHLSCHVPALLSPFGGEWVCTLC 1012
Qy 1697 LA-QQVGEF--TQKPGPKGQKRGKSGVSLNFSSEGDRRRRVLLKGRSPAAGPRYSEER 1754
Db 1013 RSLTQPEWYDCENACTNQPMWRASPLGSM-----YDQKK 1047
Qy 1755 LSPSKRRRLSRNHHSDLTFTCEIILMEMESHDAWPFKEPVNPRVLSGYRIIKNPMDFS 1814
Db 1048 -----CEKVLVSLCCNNLSLFPHEPVSP-LARHYIQLIKRPMDL 1086
Qy 1815 TMRRELLR---GGYTSSEEFADALLVPDNCQTNEDDSEVGKAGHIMRRFFESRWEFF 1871
Db 1087 IIRKLOKQDPAHYTTTPEEVVSDVRLMFWNCAKFNYPDSEVAEAGRCLEVPFGWGLKEIY 1146
Qy 1872 QGKQ 1875
Db 1147 PEKR 1150

RESULT 40
TF1A_MOUSE

ID AC Q64127; O64126; TF1A_MOUSE STANDARD; PRT; 1051 AA.
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 25-OCT-2004 (Rel. 45, Last annotation update)
DE Transcription intermediary factor 1-alpha (TF1-alpha) (Tripartite motif protein 24).
DE Name=TF1; Synonyms=TF1a, Trim24;
GN Mus musculus (Mouse).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM LONG).
RC TISSUE=Cardioma; PubMed=7744009;
RX MEDLINE=9526264; MEDLINE=9526264;
RA le Douarin B., Zechel C., Garnier J.-M., Lutz Y., Tora L., Pierrat B., Heery D., Groneneyer H., Chambon P., Losson R.;
RA "The N-terminal part of TF1, a putative mediator of the ligand-dependent activation function (AF-2) of nuclear receptors, is fused to B-raf in the oncogenic protein T18";
RT EMO J. 14:2020-2033 (1995).
RL [2]
RN SEQUENCE FROM N.A. (ISOFORM LONG).
RP STRAIN=C57BL/6; TISSUE=Brain;
RC MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RX Strausberg R.L., Feingold E.A., Grouse L.H., Dege J.G., Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D., Altshul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K., Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Haieh F., Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L., Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E., Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C., Raha S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W., Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A., Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G., Blakesley R.W., Touchman J.W., Green E.D., Dickinson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butlerfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E., Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [3]
RP INTERACTIONS WITH CBX1 AND CBX3.
RC MEDLINE=97133299; PubMed=8978696;
RX le Douarin B., Nielsen A.L., Garnier J.-M., Ichinose H., Jeanmougin F., Losson R., Chambon P.;
RA "A possible involvement of Tif1 alpha and TIF1 beta in the epigenetic control of transcription by nuclear receptors.";
RT EMO J. 15:6701-6715 (1996).
CC -!- FUNCTION: Interacts selectively in vitro with the AF2-activating domain of the estrogen receptors. Association with DNA-bound estrogen receptors requires the presence of estradiol (By similarity).
CC -!- SUBUNIT: Interacts with NR3C2 (By similarity). Interacts with CBX1 and CBX3.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=Long;
CC IsoId=Q64127-1; Sequence=Displayed;
CC Name=Short;
CC IsoId=Q64127-2; Sequence=VSP_005773;
CC -!- TISSUE SPECIFICITY: Detected in all adult tissues, with the highest expression level in testis.
CC -!- DISEASE: Participates in a chromosomal translocation that produces a Tif1a-BRAF (T18) oncogene originally isolated from a furural-induced hepatoma.
CC -!- SIMILARITY: Contains 2 B box-type zinc fingers.


```
Db 363 -----RIKEVNETNRV-----:|:|:|-----EQEIKVAIF--- 383
Qy 954 ELNGSTLIINEIDKLTESMSYRNKNWIVGRLRLKTVLAKRTGRSEVMGRPECLGR 1013
Db 384 -----TLINEINKGKSL-----:|:|:|-----LQLENTV-----K 405
Qy 1014 RRSRIMEETSGWEEBEEESIAVPGRRRRDGEVDATASSIPLEBRIQIEKL----- 1066
Db 406 ERQMKLLOQ-----:|:|:|-----QNDITGLSRQVQVHVNFTNWA 435
Qy 1067 ----SKRLOFFKLL--HSSOMLRA-----VSLGQDRYR--RYWVLPYLAGIFVGT 1112
Db 436 TASGSTALLYSKRUITTQLRHILKARCDPVAANGAIRFHCDPTFWAKNVV-----N 488
Qy 1113 EGNLVEEVIKETDLSLVAHASINPALFSMKMELAGSNNTASSPARARGPLTKPGF 1172
Db 489 LGNLVIES-----KPAQYTPNVVQVP--PGTNHLSKTPGQINLAQLRL----- 532
Qy 1173 MOPRHFKSPVRGDSQEQAOLOPBAQLHVPAPQPOLQLOLQSHKGFLEQSGSLSLGQ 1232
Db 533 ---CHMOQVVAQHQLOQ-----QMRLOQPPAPITTTATTQOH-----PRQAAPQMLQ 580
Qy 1233 ----SQHLSQSAFLSWLSQTSQSHSLSSSVLPDSSPGKLDPAQSPPEP 1281
Db 581 QPPLISVOTWQRGNMCCAF-----OAHQMRLAQNAARIPGIP--RHSAPQYSMMQP 631
Qy 1282 EPDEAESPDLOAFWNFISAQMPCN--AAPTPLPPLAVSEDOPTSPQOLA-----SSKPMN 1334
Db 632 HLQRHSPHGAGPPFVVAHNPINPTSTATTMANANRGTPSPVTAIELIPSVTNPEN 691
Qy 1335 RPSAANPCSPQVF-----SSTPLAGLAPKRAG-----DPGEMPOSP--TGLGQ 1376
Db 692 LPSLPD-IPPIOLEDAGSSLDNLLSVISGSHLPQPTSTWNPSPGPSALSPGSSGLSN 750
Qy 1377 PKRGSPSPKPKQMEQRYLTQATQVPPPEMCSGWWIIPPEMDMLKALHPRG----- 1432
Db 751 SHTPVRPES-----TSSTGS-----RGSCGS 771
Qy 1433 ---IREKALHKLHKLNRDFLQEVCLRPSADPIFEPRQLPAFOEGIMSPKERTVETDLA 1489
Db 772 SGRTAKSAHF-----KSDQV-KVQEPGTEERELCSFGAVKQKTB-- 813
Qy 1490 VLQWVEELQVRVMSDLQIRGWCPSDSTREDLAYCEHLSDSQBDITWRGREGAPQ 1549
Db 814 -----DGRSACMLSPSS-----LTFP 832
Qy 1550 RKTNPDLAVMLAALQNVKRYLREPLWT--HEVLEKALLSTPNGAPEGTTTIS 1607
Db 833 LSTNLHLESELDTLTGLENHVKE-----PTDISCKQSGLSNLVNGK----- 876
Qy 1608 YEITPRIRWTLQRCRSAHVCLGLHLERSIAWEKSVNKTCLVCRKGDNDFFLLC 1667
Db 877 -----SPINLMHRSARIQGDGNS--KDDPDNEDWCAVCQNGD--LUC 917
Qy 1668 DCDRGCHYCHRPKMEAVPEGDMFTVCYLAQQVGEFTQKPGPKQKRGKSGYSLNFS 1727
Db 918 EKCPKVFHLTCHVPTLLSPSGDMICTFC-----RDIGKE-----V 954
Qy 1728 EGDGRRRVLKGRSPAGPFRYSERISPKRRRLSMRNHHSJLTFCEIILMEMESHDA 1787
Db 955 EYDCDNMQHKKGK-----TAQGLSPVDQK-----CERLLLYLCHL 993
Qy 1788 AWPEKPEPNRIVSGYRRIIKNPMDFSTWRBLLR--GGYTSSEFAADALLVPDNCOT 1844
Db 994 SIEFOEPV-PVSIPIYKIIKKPMDSITVKKLQKKHSHQYQIPDFVADVRLIFKNCR 1052
Qy 1845 FNE-----DDSEVGKAGHIMRFFESRWEFEYQK 1874
Db 1053 FNEMKKVQVYADTQIEINLKGDSVAKAGKAVALYFEDKLSEIYSDR 1099
```

RESULT 43

```
TFIIA_HUMAN STANDARD; PRT; 1050 AA.
ID O15164; O95854;
AC 15-JUL-1999 (Rel. 38, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Transcription intermediary factor I-alpha (TIF1-alpha) (Tripartite
DE motif protein 24).
OS Homo sapiens (Human).
GN Name=TIF1; Synonyms=TFIIA, TRIM24;
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM SHORT).
RC TISSUE=Breast cancer;
RX MEDLINE=97277352; PubMed=9115274; DOI=10.1074/jbc.272.18.12062;
RA Thenot S., Henriquet C., Rochefort H., Cavailles V.;
RT "Differential interaction of nuclear receptors with the putative human
RT transcriptional coactivator hTIF1.";
RL J. Biol. Chem. 272:12062-12068(1997).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM SHORT).
RX MEDLINE=99144725; PubMed=10021217; DOI=10.1038/sj.onc.1202655;
RA Venturini L., You J., Stadler M., Gallien R., Lallemand V.,
RA Koken M.H.M., Mattei M.-G., Ganser A., Chambon P., Losson R.,
RA De The H.;
RT "TIF1gamma, a novel member of the transcriptional intermediary factor
RT 1 family.";
RL Oncogene 18:1209-1217(1999).
RN [3]
RP SEQUENCE OF 477-510 (ISOFORM LONG).
RC TISSUE=Breast cancer;
RA Cavailles V.;
RL Submitted (JAN-1999) to Swiss-Prot.
RN [4]
RP INTERACTION WITH NR3C2.
RX MEDLINE=20388588; PubMed=10935545; DOI=10.1210/me.14.8.1210;
RA Hellal-Levy C., Fagart J., Souque A., Wurtz J.-M., Moras D.,
RA Rafestin-Obelin M.-E.;
RT "Crucial role of the H11-H12 loop in stabilizing the active
RT conformation of the human mineralocorticoid receptor.";
RL Mol. Endocrinol. 14:1210-1221(2000).
CC -!- FUNCTION: Interacts selectively in vitro with the AF2-activating
CC domain of the estrogen receptors. Association with DNA-bound
CC estrogen receptors requires the presence of estradiol.
CC -!- SUBUNIT: Interacts with CBX1 and CBX3 (By similarity). Interacts
CC with NR3C2.
CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=Long;
CC IsoId=O15164-1; Sequence=Displayed;
CC Name=Short;
CC IsoId=O15164-2; Sequence=VSP_005772;
CC -!- SIMILARITY: Contains 2 B box-type zinc fingers.
CC -!- SIMILARITY: Contains 1 bromodomain.
CC -!- SIMILARITY: Contains 1 PHD-type zinc finger.
CC -!- SIMILARITY: Contains 1 RING-type zinc finger.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF009353; AAB63585.1; -
DR EMBL; AF119042; AAD17258.1; -
DR HSSP; Q14839; IMM2.
DR TRANSFAC; T04945; -
DR Genew; HGNC:11812; TIF1.
```


RA De The H.;
 RT "Tiflamma, a novel member of the transcriptional intermediary factor
 RT 1 family.";
 RL Oncogene 18:1209-1217(1999).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM ALPHA AND BETA).
 RX MEDLINE=21231161; PubMed=11331580; DOI=10.1093/emboj/20.9.2140;
 RA Raymond A., Meroni G., Fantozzi A., Merla G., Cairo S., Luzi L.,
 RA Rignelli D., Zanaria E., Messori S., Cainarca S., Guffanti A.,
 RA Minucci S., Pellicci P.G., Ballabio A.;
 RT "The tripartite motif family identifies cell compartments.";
 RL EMBO J. 20:2140-2151(2001).
 RN [3]
 RP SEQUENCE FROM N.A. (ISOFORM ALPHA).
 RN TISSUE=Brain;
 RC MEDLINE=99397452; PubMed=10470851;
 RX Kikuno R., Nagase T., Ishikawa K.-I., Hirotsawa M., Miyajima N.,
 RA Tanaka A., Kotani H., Nomura N., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. XIV.
 RT The complete sequences of 100 new cDNA clones from brain which code
 RT for large proteins in vitro.";
 RL DNA Res. 6:197-205(1999).
 RN [4]
 RP SEQUENCE OF 76-1127 FROM N.A. (ISOFORM ALPHA).
 RC TISSUE=Thyroid;
 RX MEDLINE=99367902; PubMed=10439047; DOI=10.1038/sj.onc.1202824;
 RA Klugbauer S., Rabes H.;
 RT "The transcription coactivator htif1 and a related protein are fused
 RT to the ret receptor tyrosine kinase in childhood papillary thyroid
 RT carcinomas.";
 RL Oncogene 18:4388-4393(1999).
 CC -!- FUNCTION: Seems to act as a transcriptional repressor.
 CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=Alpha;
 CC IsoId=Q9UPN9-1; Sequences=Displayed;
 CC Name=Beta;
 CC IsoId=Q9UPN9-2; Sequences=VSP 005774;
 CC -!- SIMILARITY: Contains 2 B box-type zinc fingers.
 CC -!- SIMILARITY: Contains 1 bromodomain.
 CC -!- SIMILARITY: Contains 1 PHD-type zinc finger.
 CC -!- SIMILARITY: Contains 1 RING-type zinc finger.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF119043; AAD17259.1; -;
 DR EMBL; AF220136; AAG53509.1; -;
 DR EMBL; AF220137; AAG53510.1; -;
 DR EMBL; AB029036; BAA83065.1; ALT_INIT.
 DR EMBL; AJ132948; CAB55313.1; -;
 DR HSSP; Q14839; 1M2M.
 DR TRANSFAC; T04946; -;
 DR Genew; HGNC:16290; TRIM33.
 DR MIM; 605769; -;
 DR GO; GO:0005634; C:nucleus; NAS.
 DR GO; GO:0003714; F:transcription corepressor activity; ISS.
 DR GO; GO:0008270; F:zinc ion binding; NAS.
 DR GO; GO:0016481; P:negative regulation of transcription; NAS.
 DR InterPro; IPR003649; Bbox_C.
 DR InterPro; IPR001487; Bromodomain.
 DR InterPro; IPR011011; FYVE PHD Znf.
 DR InterPro; IPR000315; Znf Bbox.
 DR InterPro; IPR001965; Znf_Phd.
 DR InterPro; IPR001841; Znf_ring.
 DR Pfam; PF00439; Bromodomain; 1.
 DR Pfam; PF00628; PHD; 1.

DR Pfam; PF00643; zf-B_box; 2.
 DR PRINTS; PR01406; BBOXZNFINGER.
 DR PRINTS; PR00503; BROMODOMAIN.
 DR SMART; SM00502; BBOX; 1.
 DR SMART; SM00336; BBOX; 2.
 DR SMART; SM00297; BBOX; 1.
 DR SMART; SM00249; PHD; 2.
 DR SMART; SM00184; RING; 2.
 DR PROSITE; PS00633; BROMODOMAIN_1; FALSE_NEG.
 DR PROSITE; PS00634; BROMODOMAIN_2; 1.
 DR PROSITE; PS00119; ZF_BBOX; 2.
 DR PROSITE; PS01359; ZF_PHD_1; 1.
 DR PROSITE; PS00016; ZF_PHD_2; 1.
 DR PROSITE; PS00518; ZF_RING_1; 1.
 DR PROSITE; PS00089; ZF_RING_2; 1.
 KW Nuclear protein; Repeat; Repressor; Transcription regulation;
 KW Zinc-finger.
 FT DOMAIN 13 18 Poly-Gly.
 FT DOMAIN 44 49 Poly-Glu.
 FT ZN_FING 125 154 RING-type.
 FT ZN_FING 212 259 B box-type 1.
 FT ZN_FING 271 312 B box-type 2.
 FT DOMAIN 299 401 Coiled coil (Potential).
 FT DOMAIN 545 550 Poly-Thr.
 FT ZN_FING 887 934 PHD-type.
 FT DOMAIN 974 1046 Bromodomain.
 FT VARSPLIC 1041 1057 Missing (in isoform Beta).
 FT CONFLICT 89 89 V -> E (in Ref. 4).
 FT CONFLICT 451 453 PAA -> LKH (in Ref. 4).
 FT CONFLICT 909 909 F -> S (in Ref. 4).
 FT CONFLICT 1037 1037 R -> T (in Ref. 1).
 FT CONFLICT 1114 1127 RKRLKSDRPVHIK -> OKTPKVR (in Ref. 1).
 SQ SEQUENCE 1127 AA; 122521 MW; 02E3880CFDCFDA3B CRC64;
 Query Match 3.4%; Score 335; DB 1; Length 1127;
 Best Local Similarity 19.5%; Pred. No. 9.5e-06;
 Matches 241; Conservative 148; Mismatches 449; Indels 400; Gaps 52;
 QY 766 PACRADKTATLORLEERQKQMIILEMKKPTEDMCLTDHQLPDPFSRVPGLTLPSPGFS 825
 DB 118 PASLLDTCAVCQSQSLRSREAE---PKLLPCLHSFCL---RCLPEPQLSVPIPGSGV 171
 QY 826 D-----CLTIVEFLHSFGVLGFDPAKDVPSVLQELGCLCG--DSLGE 868
 DB 172 DIQQGVIRCPVQCECHQIDLVNDFYFK-----DTSEAPSSSSDEKSEQVCTSCEDNAA 226
 QY 869 VODLLVRLKLAALHDPGPPSYCQS--LKILGEKVSEIPLTRD-----NVSEILRCFLM 919
 DB 227 V-----GFCVEGGEWLCKTCTEAHQVRVKFTKDHILRKEDVSESVG---- 267
 QY 920 AYGVPALCDRLRTQPPFOAQPQQAVALAPVHVLNGSTLIINEIDK-----TLESM 972
 DB 268 ASGQRPVFC-----PVHKQEQKLFCFCTCDRLTCTCDCLLEHK 305
 QY 973 -----SSYRNKNKWIVEGRLLRLKTVLAKRT-----GRSEVEMGRPEECGLRRSSRM 1020
 DB 306 EHRQFLEEFQNKGAENLLAKL-----LEKNVYHFAATQVQ-----NRK 349
 QY 1021 E--ETSGMEEEEEESIAAIVPGRRRRDEVDATASSIPELERQIEKLSKRLFFRKLL 1078
 DB 350 EWNETNKRVEQEIKAIFTL-----INEINKGKS---LLQLENVTKER---QMKLL 396
 QY 1079 HSSQMLRAVSLGQDRYRRRYWVLPYLAGIFVEGTGNLVPEE-----VIKETDSL 1129
 DB 397 QQQNDITGLS-----RQVGHVMAFNWAIASGSTALLYSKRLITFQLRHILKARCDEV 450
 QY 1130 KVAHAH---SLNPALFSMKMELAGSNNTTASPARARSPLTKPGF----- 1172
 DB 451 PAANGAIRFHCDPTFWAKNVNGLNLVIESKPA-----PGYTNVVVGVPPGTN 500
 QY 1173 -----MOPRHFKSPVRQDSQEQPAQALQPEAQLHVPAPQPOLQLQLQSHKG 1219

```
Db 501 HISKTPGQINLAQLRQHQMOQVVAQKHQQLQ-----QMRMQPPAPVPTTTTTTQOQ-- 553
Qy 1220 FLEQEGSPLSLGQ-----SQHDSQSAFLSWLSQTSQSHSLSSSSVLPDSSPG 1268
Db 554 ---PQAAQPMQLQPPQLISVOTWQRMGNCAG-----QAHQMLAQNA-----RIPG 601
Qy 1269 KLDPAQPPPEPPEDEAESPDLOAQFWFNISAQWPCNAATPPLAVSBDQTPSPQOLA 1328
Db 602 -----IPRHSQFOYSMMQPHLQHQSN-----PGHAGFPVVSVENTINTSPTTA 648
Qy 1329 SSKPMNRSAANCPSPVOFSPTPLAGLAPKRRAGDPGEMPOSPTLGQPKRGRGPP----- 1384
Db 649 TMANANR-----GPTSP-----SVTAIELIPS--VTNPENLPSL-----DIPPIQLE 689
Qy 1385 ---SKFFQMQQRVLTQTAQVPVPEMCSGHWIIPDPENLDAMLKALHPRGIREKALKHK 1441
Db 690 DAGSSSLDNLISRY---ISGSHLLPQPTSTVNPSPGFSALSPGSSGLS----- 734
Qy 1442 LNKHRDFLQVCLRPSADPIEPRLQPAFQBGIMSWSPK---EKT---YETDLAVLQWVEE 1496
Db 735 -NSH-----TPVRPPSTSTGSGSGSSGRTAEKISLSPKSOQVKVQEPG 780
Qy 1497 LEQRVIMSDLOIRGWTCPSPDSTREDLAYCEHLSDSQSDITWRGFGREGLAPQRKTTNPL 1556
Db 781 TEDEI-----CSFGGVKQ-----EKTEDGRSACWLSPPSSSLTPPLSTNLHL 824
Qy 1557 DLAVWRLAALQONVKRRYLRLPETHVLEKALLSTPNQAPAGPTTTEIYEITPRIRI 1616
Db 825 ESELDALASLHNVK---TEPA--DMNESCQSGLSLNVNGK----- 861
Qy 1617 WRQTLQRCRSHAVCLCLGHILERSIAWEKSNVKTCLVCRGDNDEPILLCDGDRGCHI 1676
Db 862 -----SPIRSLMHRSAARTGGDNN--KDDDPNEQWCAVQONGD---LLCCEKCPKVPHL 911
Qy 1677 YCHRPQMEAVPEGDFWCTVCLAQAVEGEFTQKPGPKRQKRKSGYSLNFSEGDGRRRV 1736
Db 912 TCHVPTLLSPSGDWICTFC----- 931
Qy 1737 LLKGRSPAAGPRYSEERLSPSKRRRLSMRNNHSDLTFCETILLMEMESHDAAMPFXEPVN 1796
Db 932 ---RDIGKPEVEDCNLOHKKKGTAAQGLSPVDQRKCEKRLLYLCHELSLBQEPV- 986
Qy 1797 PRLVSGVRRITKNPDMFSTMERLLR---GGYTSEBEAADAALLVDFNCOTFNE----- 1847
Db 987 PASIPNYKIITKPMWDLSTVKKLQKHSQHYIPDDFVADVRLIFKNCERFNEWMKVQ 1046
Qy 1848 -----DDSEVGKAGHIMRRFFRSRWEFFYQKG 1874
Db 1047 VYADTQELNKADSEVAQAGKAVALLYFDKLTETIYSDR 1084

RESULT 45
Q7SDK2 PRELIMINARY; PRT; 10820 AA.
AC Q7SDK2
DT 01-MAR-2004 (TReMBLrel. 26, Created)
DT 01-MAR-2004 (TReMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TReMBLrel. 26, Last annotation update)
DE Hypothetical protein.
GN Name=NCU02793.1;
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OR74A;
RA Galagan J.E., Calvo S.E., Borkovich K.A., Selker E.U., Read N.D.,
RA Jaffe D., FitzHugh W., Ma L.-J., Smirnov S., Purcell S., Rehan B.,
RA Elkins T., Engels R., Wang S., Nielsen C.B., Butler J., Endrizzi M.,
RA Qui D., Iankiev P., Pedersen D., Nelson M., Washburne M.,
RA Selitrennikoff C.P., Kinsey J.A., Braun E.L., Zelter A., Schulte U.,
```

```
RA Kothe G.O., Jedd G., Mewes W., Staben C., Marcotte E., Greenberg D.,
RA Roy A., Foley K., Naylor J., Thomann N., Barrett R., Gnerre S.,
RA Kanal M., Kanvasselis M., Mauceli E., Bielek C., Rudd S., Frishman D.,
RA Krystofova S., Rasmussen C., Metznerberg R.L., Perkins D.D., Kroken S.,
RA Cogoni C., Macino G., Catchside D., Li W., Pratt R.J., Osmani S.A.,
RA DeSouza C.C., Glass L., Orbach M.J., Berglund J., Voelker R.,
RA Yarden O., Plamann M., Sella J., Dunlap J., Radford A., Aramayo R.,
RA Natvig D.O., Alex L.A., Mannheim G., Ebbole D.J., Freitag M.,
RA Paulsen I., Sachs M.S., Lander E.S., Nussbaum C., Birren B.,
RA "The Genome Sequence of the Filamentous Fungus Neurospora crassa.";
RL Nature 0:0-0(2003).
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMEL; AABX01000050; EAA34839.1; -.
DR HSSP; P01096; IGMJ.
KW Hypothetical protein.
SQ SEQUENCE 10820 AA; 1175573 MW; C5F7D4FC763FB1BF CRC64;

Query Match 3.3%; Score 320; DB 2; Length 10820;
Best Local Similarity 21.0%; Pred. No. 0.0063;
Matches 368; Conservative 231; Mismatches 613; Indels 544; Gaps 89;

Qy 17 LPPAPAAAGLKPSPSSGEGLYTNGSPMNFPOQKSLNGDVNVNGLSTVSHTTTSGILNSA 76
Db 6424 MPPPPEDIELPPSPDQVE-----HREKPIVAIEIHERVKSPVSGHQDVETPSR 6473
Qy 77 PHSSSTSHLHPHSVAYDCLWNYSOYPSANPGSNLKPPLLSQFGSGQYPLNGILGGSQRP 136
Db 6474 EHSPA-----HSPAVEV-----SKSEVPTAAAEPEVEEDL----- 6505
Qy 137 SSPSHNTNLRAGSKQFWANGTHSPMGLNFOSQELYDSFPDQNFEEVCSGIHP---DEAAE 193
Db 6506 ASQSKQKTLAELEEFQSGSPSPPEVDNRDAELV--SFTGLTKDAELSSAHSHPMDAPTD 6563
Qy 194 KEMTSVVAENGTGLVCSLELEEXQPELKMCGYNSVPSVESLHDEV--SVLVP---DPTVS 249
Db 6564 VELPA-----SPSPKETDDQDKDLPTTTSHEAIAINVITPTWTSPTVS 6607
Qy 250 CLDDPSHLPOLETP-----ILSEDSLEFPNSLAPPEVSGGLYIGDDTEL-MGAEDKL 302
Db 6608 ---PKVKPENIELPPSPQRPQLLPSNIVPGYQFSPEPIS-----DDERLPLSPDDEF 6657
Qy 303 PLXDSPIVIALDCSLNNAATAFLAADSOTSTISIFASPTSPPVLGESVLQDNFSLNNG 362
Db 6658 EDDDDGSGPVDVQYERS-----FSDDPOEG--FAS--SP----- 6690
Qy 363 SDAEQEMETQSDFPPLSLTOPAPDQSQSTIQLHPA-----TSPAVSPPTSPAVSLVSPA 417
Db 6691 DEATVDHVEVQHS--PFSQSGPA--FASSPRISPAQVELPASVESPNFAPA-----SSS 6742
Qy 418 ASPEISPEVCPAASTVVSPAV-----FSVSPASSAV----- 449
Db 6743 PSPKSTPE----DAVLSPALEAAAEFTPSPSPVSPSPSTKDTTWMFERGVSSASASAPA 6797
Qy 450 -----LPAVSLEV--PLTASVTS-----PKAS-----PVTSPAAAPP-- 479
Db 6798 PAPEENVQFAPAPSPSEVQHPASPCKISSEDVGLPKSLLEPEAEHEMELPTIAAAQALPED 6857
Qy 480 ---TASPANKDVSSFLETTADVEE--ITGEGLTASGSDVMRRRIATPEEVRLPLQHCWR 534
Db 6858 VELPMSFA-----ETEPKHELPLITKEIPLTAHELPTSLPKI-SPEEVELPMS----- 6904
Qy 535 REVRIKKGSHRWQGETWYGFQGRMKQKQFPEVIKYL--SRNLVHSVRREHFSFGPRMPVG 592
Db 6905 PEVEVPTPAH-----EVIQSLGQSDNIHKV-----SSSPKIAPA 6939
Qy 593 DFFEERTPEGLQWQLSABEIPRIQAITKGRGRPRNTEKAKTEKVEPK-VKRGGRGPPK 651
Db 6940 DI-----ELPT-----SPEN-----NVLDSFKTLPLPGDSSLPE 6967
Qy 652 VKITELLNKTDNRPLKLEAQETLNEE---DKAKIAKSKKQKQKQVORGECLTT---IQG 705
```

Db 6968 SHLTAQVEKDWDLSTAAEQDDQIYVDEQSDNDLEVSTSHKIVPEVEVLPSPSPVPVEG 7027

Qy 706 QARNKRKQETSLKHKEAKKSXAEBKKGTKQEKLEKVKREKKEKVKMEKEEVTAK 765

Db 7028 -----LDVNEIQHLHSQIIPHVQRADSTWGEAELEVPSPSPLEIAHKEDELPA 7080

Qy 766 PACKADKTLATORRLEERKQKQMLLEMKKPTEDMCLTDHQPPLDFSRVPGTLPSGAPS 825

Db 7081 PV-----KEAAYEE-----VEVPASPVKAAAKD-----IEVQKSPSPAAE 7116

Qy 826 DCLTIVEFLHSGKVLGFDPAKDV-----SLGLVQELGLCCQDSLGEVQDILLRLK 878

Db 7117 LDVTEIQHLHS-----SQIIPHVQRAVSPAVIHEEV-----QSTPLDVPVEVEL-- 7163

Qy 879 AALHDPGPPSYCQSLKILGEKVEIPLTRDNVSEILRCFLMAYGVXPALCDRL-----RTOP 935

Db 7164 -----PASPVKEAALEDIEALRSPSPAABLDVTEIQH-----LHSSQILPHIERAVSSNTVP 7216

Qy 936 FQAQPPQKAAVLA-----FPVHELNGSTLIINEIDKTLSESSSYRKNKWIIEGELRLK 990

Db 7217 EDVKAPASPAETAPKEDVLPSPSPGAELDVTEIQHLHSQ-----ILPHIQRAD 7268

Qy 991 TVLAKTGRSEVMGRPECLGRRRSRIMBETSGEEREEESTAAVPGRRRRGRDGV 1050

Db 7269 STWGE-VEPEVELPKPP-----VEEIVHEE-----VHVPAFKVVGSEVE 7308

Qy 1051 ATASS-----IPELEROIEKLSKQLFRKLLHSSQMLRAVSLGQDQRYRRY 1098

Db 7309 NPASSPSPKIAEEMDIPRSPSPVDNDVTELQHQ-----HSTQI----- 7349

Qy 1099 WLPVLAGIFVEGTGNLVPEVVIKETTDSLKVAHASLNPAFLSKMELAGSNNTASSP 1158

Db 7350 -IPHTQSVLPFTAQVAPPEI-----QLPASPLEQFTPE-----EVELPRS----- 7391

Qy 1159 ARASRLPKTKPGFQMPRHFS-----PVRGDSQEQQAQLOPE-----AQL 1200

Db 7392 -----SSPVEAKLDIVMEQHEISAQLPLYVRADSTWGEVDPEDVDLTGTLKEIMPAEA 7447

Qy 1201 HVPAPQPP-----QLQLQLOSHKGFLEQSGPLS-----LGOSQHDLSOSA 1241

Db 7448 ELPGSPSPVEDELQAKLQ-----AAHEEVELQASEPTSPIKITEIEVPRSPSPFDESKIT 7504

Qy 1242 FLSWLSQTSQSHSSLLSSSVLPDSSPG--KLDP-----APSQPPEPEPEDEAESPDLOAP 1295

Db 7505 ELQHLHSAQ-----IIPRNHYATSSSGSPKIIEVEVLPASPVVEVDVDE-----LQRF 7554

Qy 1296 WFN--ISAQPCNAAPTPLAVSEQPTP-SP-----QQI-----ASSKPMNRPSAANPCSP 1344

Db 7555 HSSQIISPHIERSASPTIEEVVPEVEVLPSPVEQVQVIEQLOHLHSROIIPPHIQHPASP 7614

Qy 1345 VQFSSTPLAG-LAPKRRAGDQEMQPSPTGLQPKRRGRPPSPKPKQEQRYLTOLTAAQP 1403

Db 7615 -----SPNAELLASPKVTPPEEVELPASPELVPQDIELPASPIDEAQELPILKESEAS 7669

Qy 1404 VPPENCSGWNIIPD-PEMLDAMLKALHPRGIREKALHKLHNKGRDFLOEVCLRPSADIP 1462

Db 7670 VDQO-----IDDLLEEDARSAALELPSPEV-----KHEVQPEEMELPLS--BVL 7713

Qy 1463 EPRQ--LPAFQEGINSWSPKETYETDLAVLOWVEELQRYVMSDLOIRGWCSPSPDSTR 1520

Db 7714 KPVEVDLPA-----SPVEVPEEVELFVSP-VDDL-----ADVQ-----HLPODDVP 7754

Qy 1521 EDLACEHLSDSQEDITWRGPGREGIAPQRTTNPLDLAVMLAALQNKRRYLRPLW 1580

Db 7755 EAV-----QENVNLEIPSSPIAPEE-----VELPLSKTAPAE-----VDVPAS 7795

Qy 1581 PTHEVLEKALLSTPN 1596

Db 7796 PWVEVEEQPATVSEPS 7811

RESULT 46
Q6E2N2

ID Q6E2N2 PRELIMINARY; PRT; 961 AA.

AC Q6E2N2;

DT 25-OCT-2004 (TREMBlrel. 28, Created)

DT 25-OCT-2004 (TREMBlrel. 28, Last sequence update)

DT 25-OCT-2004 (TREMBlrel. 28, Last annotation update)

DE Transcriptional intermediary factor 1 alpha

OS Brachydanio rerio (Zebrafish) (Danio rerio).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;

OC Cyprinidae; Danio.

OX NCBI_TaxID=7955;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Kidney marrow;

RA Ransom D.G., Bahary N., Niss K., Traver D., Burns C., Trede N.S.,

RA Palfett-Lugassy N., Saganley W.J., Lim C.A., Hersey C., Zhou Y.,

RA Barut B.A., Lin S., Kingsley P.D., Palis J., Orkin S.H., Zon L.I.;

RT "The zebrafish moonshine gene encodes transcriptional intermediary

RT factor 1 gamma, an essential regulator of hematopoiesis.";

RL PLoS Biol. 2:e237-e237(2004).

CC -!- SIMILARITY: Contains 1 RING-type zinc finger.

DR EMBL; AY598454; AAT70733.1; -

DR GO; GO:000151; C:ubiquitin ligase complex; IEA.

DR GO; GO:0003677; F:DNA binding; IEA.

DR GO; GO:0004842; F:ubiquitin-protein ligase activity; IEA.

DR GO; GO:0008270; F:zinc ion binding; IEA.

DR GO; GO:0016567; P:protein ubiquitination; IEA.

DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.

DR InterPro; IPR003649; Bbox C.

DR InterPro; IPR001487; Bromodomain

DR InterPro; IPR011011; FIVE_PHD_Znf.

DR InterPro; IPR000315; Znf_Bbox.

DR InterPro; IPR001965; Znf_PHD.

DR InterPro; IPR001841; Znf_ring.

DR Pfam; PF00439; Bromodomain; 1.

DR Pfam; PF00628; PHD; 1.

DR Pfam; PF00643; zf-B_box; 2.

DR Pfam; PF00097; zf-C3HC4; 1.

DR PRINTS; PR01406; BBOXZNFINGER.

DR PRINTS; PR00503; BROMODOMAIN.

DR SMART; SM00502; BBC; 1.

DR SMART; SM00336; BBOX; 2.

DR SMART; SM00297; PHD; 1.

DR SMART; SM00249; BROMO; 1.

DR SMART; SM00184; RING; 1.

DR PROSITE; PS00014; BROMODOMAIN_2; 1.

DR PROSITE; PS00119; ZF_BBOX; 2.

DR PROSITE; PS01359; ZF_PHD_1; 1.

DR PROSITE; PS00016; ZF_PHD_2; 1.

DR PROSITE; PS00518; ZF_RING_1; 1.

DR PROSITE; PS00089; ZF_RING_2; 1.

DR Metal-binding; Zinc; Zinc-finger.

QW SEQUENCE 961 AA; 107346 MW; 0C768A7E03ED64BB CRC64;

Query Match 3.2%; Score 312.5; DB 2; Length 961;

Best Local Similarity 20.7%; Pred. No. 6.6e-05;

Matches 161; Conservative 84; Mismatches 238; Indels 295; Gaps 31;

Qy 1117 VPVEVVKETOSLKVAAHASLNPAFLSKMELAGSNNTASSPARASRLPKTKPGMOPR 1176

Db 370 VPQSLVRFQCSAFWASNVLD-----GSLVVEKTPVRHPSGP-----QPF 409

Qy 1177 HFKSVPVRQDSQEQQAQLOPEAQLHVPQAPQLOQLQSHKGFLEQSGPLSLQSQSHD 1236

Db 410 PFOQMNFPRAETLQOQQOQOQOQO-----STLAQLQMOVEK-----LSQHN 449

Qy 1237 LSQSAFLSWLSQTSQSHSSLLSSSVLTPDS-----SPGLDPAQSPPEPEPEDEAESPDLO 1293

Db 450 -----SRHS-----PNQWTWYPGMRLFGP-PPPHRLQGGSPS----- 482

Qy 1294 AFWFNISAQPCN-----AAPTPLAVSEQPTPSPQOLASSKPMNRPSAANPCSPVQFSS 1349

Db 483 -----QTLPMNPHHGRYRGARSPP-----SMLQPSNLPPTQTLRGL-INNPT--YPPKPVVEVLP 532

```
QY 1350 TPLAGLAPKRRAGDPGEMQPSQPTGLGPKRRGRPPSKFFKQMEQRYLTQLTAQPVPPBMC 1409
DB 533 ---RGRFQAAAYSGGGLPSPNTLQQNQMA-----TSYLNK-----RC 569
QY 1410 SGWWIPDPMDLAKLHPRGIREKALHKLHN-----KHRDFLQEVCLRPSADPFI 1462
DB 570 -----DP-----SALMSAQR-----SYTHLSMPTETAAGQRFQSMV--TSSEP-- 610
QY 1463 EPROLPAFQEGIMSNPKPKYETDVLAVLQWVEELQORVIMSDLIQIRGWTCPSPDSTRED 1522
DB 611 -----KTGSVSWKRAE----- 621
QY 1523 LAYCEHLSDSQEDITWRGPGREGLAPORKTTPNPLDLAVMLRAALEQNVKRYLREPLWPT 1582
DB 622 -----APOSQSPN-----STKRRRRSP----- 640
QY 1583 HEVLEKALLSTPNAGPEG-----TTTEISYEITPRIRWQTLQRCRSAAHVCLCLGH 1636
DB 641 GPVIVIK-----DEPEDDDEVPRVNTSASASLPDSTGVQSQVQGEQQS----- 685
QY 1637 LERSIAWKSNNVTCLVCRKGNDEFLLLCGDGRGCHIIYCHRPKMEAVPEGDFWCTVC 1696
DB 686 -EKTPEADEDNEDWCAVCONGGE---LICCCKCPKVHLSCHVPSLTASPSGEWYCTLC 741
QY 1697 LAQVGEGETQKPGPKRGQKRSYLNFSFGDGRRRRVLLKGRSPAAGPRYSEERLS 1756
DB 742 -----RDLSNPMQYNNVAGG 757
QY 1757 PSKRRRLSMNRH---HSLDTFCEILMEMESHDAAPXEPVNPRLVSGYRRIKPNMDF 1813
DB 758 ESKDLQDLHSETHTVDRRCERLLLYCNEIUSDFOEITPSSMPEYSEIKTPMDL 817
QY 1814 STMERLLRGYTSSEFAADALVDFNCOTFNEDSDSVGKAGHMRFFPSRWEEFY 1871
DB 818 SWRSKLEDSQYKSTEDFVADVRLIFKNCATFHKEDTMSVGVANLESFFPEEQKLKLY 875

RESULT 47
Q95T46 PRELIMINARY; PRT; 249 AA.
AC Q95T46; Q81MD4;
DT 01-DEC-2001 (TremBLrel. 19, Created)
DT 01-DEC-2001 (TremBLrel. 19, Last sequence update)
DT 25-OCT-2004 (TremBLrel. 28, Last annotation update)
DE GH22615p (CG10897-P8).
GN Name=tou; ORFNames=CG10897;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA Nuncio J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,
RA Yu C., Lewis S.E., Rubin G.M., Celniker S.;
RA Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
RA Abril J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
```

```
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foaier C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong P., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattai B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.B.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA Williams S.M., Woodage, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RA "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426065; PubMed=12537568;
RA Celniker S.E., Wheeler D.A., Kronmiller B., Carlson J.W., Halpern A.,
RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,
RA George R.A., Hoskins R.A., Lavery T., Muzny D.M., Nelson C.R.,
RA Pacleb J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,
RA Svirskas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
RA Weinstock G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.;
RA "Finishing a whole-genome shotgun: Release 3 of the Drosophila
RA melanogaster euchromatic genome sequence."
RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426070; PubMed=12537573;
RA Kaminker J.S., Bergman C.M., Kronmiller B., Carlson J., Svirskas R.,
RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,
RA Ashburner M., Celniker S.E.;
RA "The transposable elements of the Drosophila melanogaster euchromatin:
RA a genomes perspective."
RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).
RN [5]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426069; PubMed=12537572;
RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,
RA Smith C.D., Rupy J.L., Whitfield B.J., Bayraktaroglu L., Berman B.P.,
RA Bettencourt B.R., Celniker S.E., de Grey A.D., Drysdale R.A.,
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA Lewis S.E.;
RA "Annotation of the Drosophila melanogaster euchromatic genome: a
RA systematic review."
RL Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).
RN [6]
RP SEQUENCE FROM N.A.
RG FlyBase;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RN [7]
RP SEQUENCE FROM N.A.
RG FlyBase;
RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY060335; AAL25374.1; -.
DR EMBL; AE003825; AAM68713.3; -.
```


HSPF; Q03330; 156I.
DR FlyBase; FBgn0033636; tou.
DR InterPro; IPR001487; Bromodomain.
DR Pfam; PF00439; Bromodomain.1.
DR PRINTS; PR00503; BROMODOMAIN.
DR SMART; SM00297; BRMO; 1.
DR PROSITE; PS00014; BROMODOMAIN_2; 1.
SQ SEQUENCE 249 AA; 27530 MW; 01DE78DF0955890 CRC64;

Query Match 3.1%; Score 307; DB 2; Length 249;
Best Local Similarity 47.1%; Pred. No. 2.le-05;
Matches 56; Conservative
 0; Indels 0; Gaps 0;

Qy 1751 SEERLSPSKRRIRSRNHHSDLTCEIILMEMESHDAAWPFEPVNPRLVSGYRRIKNP 1810
 ::|: |::: :|: |::: :|: |::: :|: |::: :|: |::: :|: |::: :|: |::: :|: |:
Db 125 AKEKLQKQEKKHATKKMLKELAVCKTLGEMELHDSWPELLPVNTKQPTTKKIITP 184

Qy 1811 MDESTWEERLLRGYTSSSEFAADALLAVFNCQTFFNEDDSEVGKAGHIMRRFPESRWEE 1869
 |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:
Db 185 MDLSITKKKLQDLQSYKTREDPCVDVRQIFDNCENEDDSVPVKAGHGMRKFPESRWGE 243

RESULT 48

Q8MKZ9 PRELIMINARY; PRT; 460 AA.

AC Q8MKZ9;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE CG10897-PD.

GN Name=tou; ORFNames=CG10897;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]

SEQUENCE FROM N.A.

RP MEDLINE=20196006; PubMed=107311132; DOI=10.1126/science.287.5461.2185;
RX Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Anantides P.G., Scher S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gaber G.L.,
RA Abril J.F., Agbayani A., An H.J., Andrews-Pfankoch C., Baldwin D.,
RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bernan B.P., Shandari D., Bolstein S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foeller C., Gabriellian A.E., Garh N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harritis N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwan C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Keichum K.A.,
RA Kimmei B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laško P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Paclet J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert C., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun B.,
RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodagef, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,

```
QY 1811 MDFSTMRLRLGGVTSSEEPADALLVFDNCOTFNEDDSEVGKAGHIMRRFFESRWEE 1869
DB 396 MDLSTIKKQLQDLSYKTRDFCVVRQIFDNCMFNEDDSPVGKAGHMRKFFESRWGE 454

RESULT 49
Q8R154 PRELIMINARY; PRT; 516 AA.
AC Q8R154;
DT 01-JUN-2002 (TremBLrel. 21, Created)
DT 01-JUN-2002 (TremBLrel. 21, Last sequence update)
DE Trim24 protein (Fragment).
GN Name=Trim24;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A. TISSUE=Mammary tumor;
RC STRAIN=mix FVB/N; TISSUE=Mammary tumor;
RX MEDLINE=2238257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J.J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Brownstein M.J., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek A.M., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettunen M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Griewood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krywinski M.I., Skalska U., Smallus D.E., Schmerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RL and mouse cDNA sequences.";
RL proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=mix FVB/N; TISSUE=Mammary tumor;
RA Strausberg R.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC025482; AAH25482.1; -.
DR HSSP; Q14839; 1MW3.
DR MGB; MGI:109275; Trim24.
DR GO; GO:0005737; C:cytoplasm; IDA.
DR InterPro; IPR001487; Bromodomain.
DR InterPro; IPR011011; FYVE PHD_ZnF.
DR InterPro; IPR001965; ZnF_PHD.
DR Pfam; PF00439; Bromodomain; 1.
DR Pfam; PF0628; PHD; 1.
DR PRINTS; PR00503; BROMODOMAIN.
DR SMART; SM00297; BROMO; 1.
DR SMART; SM00249; PHD; 1.
DR PROSITE; PS00014; BROMODOMAIN 2; 1.
DR PROSITE; PS01359; ZF_PHD 1; 1.
DR PROSITE; PS00016; ZF_PHD_2; 1.
FT NON TER 1
SQ SEQUENCE 516 AA; 57113 MW; 3B20FB95ED64F299 CRC64;

Query Match 3.1%; Score 307; DB 2; Length 516;
Best Local Similarity 22.1%; Pred. No. 5.2e-05;
Matches 155; Conservative 61; Mismatches 171; Indels 314; Gaps 32;

QY 1237 LSQAFLSW-LSQTQSHSLSSSLVLPDSSP-----GKLDPAF-----SOP----- 1277
DB 27 LAQQAIKQWQISSVQAPPTTSSSSSTP--SPTTISAAGYDGKAFSSPMIDLSAPVGGSY 85
```

RA Fahy J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzyzinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Brain;
RA Strausberg R.;
RL Submitted (NOV-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC060715; AAH60715.1; -
DR GO; GO:0003677; P:DNA binding; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR001487; Bromodomain.
DR InterPro; IPR006209; EGF-like.
DR InterPro; IPR001965; ZnF_PHD.
DR Pfam; PF00439; Bromodomain; 1.
DR Pfam; PF00628; PHD; 1.
DR PRINTS; PRO0503; BROMODOMAIN.
DR SMART; SM00297; BROMO; 1.
DR SMART; SM00249; PHD; 1.
DR PROSITE; PS00633; BROMODOMAIN_1; 1.
DR PROSITE; PS00014; BROMODOMAIN_2; 1.
DR PROSITE; PS01186; EGF_2; UNKNOWN_1.
DR PROSITE; PS01359; ZF_PHD_1; UNKNOWN_1.
DR PROSITE; PS00016; ZF_PHD_2; 1.
SQ SEQUENCE 1114 AA; 121422 MW; 7A3906159D59599B CRC64;

Query Match 3.1%; Score 301.5; DB 2; Length 1114;
Best Local Similarity 19.5%; Pred. No. 0.00022;
Matches 174; Conservative 125; Mismatches 333; Indels 261; Gaps 33;

QY 1152 NTTASSPARARSRPLK-----TKGFMQPRHFKSP----- 1181
DB 308 STAISAPSTASSAPVQKGLTPGAAGPLQPSAPHSPRQGGQVQLTMAQLTLTQHGHN 367

QY 1182 -----VRGDSQEQAOLOPEAQLHVAQPOLOLOLQSH--KGFLEQSGSLSLQCS 1233
DB 368 QGLTVVIGQGQTTGOLQILPGQMTVLPGPGQQLMQAAMPNGTVQRF-----FTPLSTSAT 424

QY 1234 QHDLQSQAFLWSLT-----QSHSSLSSSLSVLPDS-----SPGKLDPAQSP 1277
DB 425 AASSSSNSSTTTNATAAGSGEQKSKILPQTQVPATTLAPTQSSSVSPAQAQPAQ 484

QY 1278 PEEPEPD-EAESSPLQAFWFNISAQPCNAAPTPLAVSEDPPTPS-PQQLASKPMNR 1335
DB 485 AAQPOQPPPAQPEVQ-----TQP-AVSSHVPSETQPSQAQTSKPL-- 525

QY 1336 PSAANPC-----SPVQSSSTPLAGLAPKRRAG-DPEMPQ-----SPTG 1373
DB 526 --VATQCPQSSVQSGSPVRVQSPPLTRPSTPSQVTPGQGPVQVTTASQPIPIPPPTS 583

QY 1374 L-----GPKRRGRPPSKFFKQMEORYLTOLTA-----QPVPEMCSGWWIPDPEMLD 1422
DB 584 LQAPSGQGPQSPQVQSSTQTLSSCQTLNQVTLVSPSPCPQPOQVIA-----VPOLQVQ 638

QY 1423 -----AMLKALH---PRGIR-EKALHKHLNKRHDFLQEVCLRPSADPIFEPRQLP 1468
DB 639 VLSQISQSVVIAQIQAGQSGVPQIKQLPIQVQNSAAQTSQSVTVQAAS-----VQEQIQ 694

QY 1469 APQEGIMSWSPKEKTYETD-----LAVLOWVELEQEVIM-----SDLIQIGWTCPSPDS 1518
DB 695 RVQQLRDQOQKKQKIETEREHTLQASNQSEIIQKQVVMKGNVIEHLKQKKTWTPAERE 754

QY 1519 TREDLAYCEHLS-----DSQEDITWRGPGREGAPOR-----KTTNPLD 1557
DB 755 ENQRMIVCNQVMKYILDKIDKEEKQAARKREBSVEQKRSQNASKUSALLFRHKEQLK 814

QY 1558 LAVMRLAA-----LEQNVKRRYLREPLWPTHEVVLEKALLSTPN----- 1596

DB 815 AEILRKRALDLKELQIQVQBELKRDLLKMKRERE---MAQAVQANAASVPTPSVPAPVAP 871
QY 1597 -----GAPEGTTTTSISYEITPR-----IRIWRQTLQRCRSAAHVCLCLGHLERSIAWEK 1645
DB 872 APAAPPAPRSPPTSTHSLPPAGHPTAPLPVTSQKRKEEEKDSKSKKKMISTTSKEAK 931
QY 1646 SVNKVTCLVCKKGNDBFLLLCDGCDRGCHYYC-HRPKMEAVPBGDWPFCTVCLAQOQVEGE 1704
DB 932 KDTRELYCICKTPYDESKFYIGCDRCQNWYHRCVGIILQSEADLIDEVVCPQC--QSTEDA 989
QY 1705 FTQKPGFPKQCKKSGVSLNFSGDRRRRLVKGRESPAAGPRYSERLSPSKRRLS 1764
DB 990 MTV-----LTPLTEK--- 999
QY 1765 MRNHSDLTFCBIILMEMESHDAAPFEPVNPRLVSGYRRIIKNPMDFSTMRLRLRG 1824
DB 1000 -----DYEGLKRVLRSLQAHKMAWPFLEPVPDNDADPYGVIVKEPMDLATWEERIQKEY 1053
QY 1825 YTSSEFAADALLVDFNCOTFNEDDSEVGKAGHIMRRFFESRWEEFYOGKQAN 1877
DB 1054 YEKLTFFVADMTKIFDNCRYNPRDTPFYQCAEVLESFFVQKLGFKASRSHN 1106

Search completed: March 9, 2005, 14:20:00
Job time : 259.909 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 9, 2005, 13:54:12 ; Search time 26.9379 Seconds
(without alignments)
4242.638 Million cell updates/sec

Title: US-10-702-148-29
Perfect score: 7985
Sequence: 1 MAPLGRKPPFLVNPFGSE.....NCFMMLVNTQFCALTDVT 1531

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 65 summaries

Database : Issued Patents AA:*
1: /cgn2_6/prodata/1/aa/5A COMB.pep:*
2: /cgn2_6/prodata/1/aa/5B COMB.pep:*
3: /cgn2_6/prodata/1/aa/6A COMB.pep:*
4: /cgn2_6/prodata/1/aa/6B COMB.pep:*
5: /cgn2_6/prodata/1/aa/PCTUS COMB.pep:*
6: /cgn2_6/prodata/1/aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	7985	100.0	1531	US-09-418-710-29	Sequence 29, Appl
2	7985	100.0	1531	US-09-839-479-29	Sequence 29, Appl
3	7955	99.6	1527	US-09-418-710-27	Sequence 27, Appl
4	7955	99.6	1527	US-09-839-479-27	Sequence 27, Appl
5	7876	98.6	1525	US-09-418-710-69	Sequence 69, Appl
6	7876	98.6	1525	US-09-839-479-68	Sequence 68, Appl
7	7662	96.0	1540	US-09-949-016-7037	Sequence 7037, Ap
8	930.5	11.7	1674	US-09-418-710-1	Sequence 1, Appl
9	930.5	11.7	1674	US-09-839-479-1	Sequence 1, Appl
10	919	11.5	1673	US-09-418-710-70	Sequence 70, Appl
11	919	11.5	1673	US-09-839-479-69	Sequence 69, Appl
12	637	8.0	128	US-09-513-999C-7706	Sequence 7706, Ap
13	533	6.7	1372	US-09-418-710-21	Sequence 21, Appl
14	533	6.7	1372	US-09-839-479-21	Sequence 21, Appl
15	518.5	6.5	1876	US-09-418-710-71	Sequence 71, Appl
16	518.5	6.5	1876	US-09-839-479-70	Sequence 70, Appl
17	512.5	6.4	1878	US-09-418-710-13	Sequence 13, Appl
18	512.5	6.4	1878	US-09-839-479-13	Sequence 13, Appl
19	509.5	6.4	1869	US-09-418-710-72	Sequence 72, Appl
20	509.5	6.4	1869	US-09-839-479-71	Sequence 71, Appl
21	352	4.4	60	US-09-513-999C-7181	Sequence 7181, Ap
22	311.5	3.9	800	US-09-270-767-45282	Sequence 45282, A
23	311	3.9	65	US-09-418-710-68	Sequence 68, Appl
24	311	3.9	65	US-09-839-479-67	Sequence 67, Appl
25	302.5	3.8	2733	US-09-949-016-11433	Sequence 11433, A
26	302.5	3.8	3259	US-09-949-016-6507	Sequence 6507, Ap
27	299	3.7	2375	US-09-538-092-1131	Sequence 1131, Ap

RESULT 1
US-09-418-710-29
; Sequence 29, Application US/09418710
; Patent No. 6596482
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042001
; CURRENT APPLICATION NUMBER: US/09/418,710
; CURRENT FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 29
; LENGTH: 1531
; TYPE: PXT
; ORGANISM: Homo sapiens
US-09-418-710-29

ALIGNMENTS

Query Match 100.0%; Score 7985; DB 4; Length 1531;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1531; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MAPLGRKPPFLVNPFGSEPFPTTPTQEAFTREYEARLYSERIWTCKTSSQL 60
|||||

Sequence 3, Appli
Sequence 7, Appli
Sequence 32, Appli
Sequence 91, Appli
Sequence 1084, Ap
Sequence 7111, Ap
Sequence 7112, Ap
Sequence 7113, Ap
Sequence 31, Appli
Sequence 4, Appli
Sequence 1252, Ap
Sequence 1246, Ap
Sequence 94, Appli
Sequence 1280, Ap
Sequence 7659, Ap
Sequence 7660, Ap
Sequence 1154, Ap
Sequence 1, Appli
Sequence 67, Appli
Sequence 66, Appli
Sequence 6, Appli
Sequence 1078, Ap
Sequence 2, Appli
Sequence 914, App
Sequence 1044, Ap
Sequence 1, Appli
Sequence 917, App
Sequence 1077, Ap
Sequence 8134, Ap
Sequence 10872, A
Sequence 1, Appli
Sequence 4, Appli
Sequence 4, Appli
Sequence 4, Appli
Sequence 1, Appli

Db 1 MAPLLGRKPPFLVNPPLGPEPFFTIPTHTQEAFTREEYEARELRYSERIWTCKTSSQOL 60
QY 61 THKEAWEEOQVAELLKEEFPWYKLVLEVMVHNTASLEKLVDTAWLEIMTKYAVGSEC 120
Db 61 THKEAWEEOQVAELLKEEFPWYKLVLEVMVHNTASLEKLVDTAWLEIMTKYAVGSEC 120
QY 121 DFEVGKEMKLVKIKIHPLEKVDDEATEKSGDGCADSPSSDKENSSOIAQDHQKKTWV 180
Db 121 DFEVGKEMKLVKIKIHPLEKVDDEATEKSGDGCADSPSSDKENSSOIAQDHQKKTWV 180
QY 181 KEDEGRRESINDRARRSPKLTSLKGERKWAPPKFLPHKYDVKLQNEDKIIISNPADS 240
Db 181 KEDEGRRESINDRARRSPKLTSLKGERKWAPPKFLPHKYDVKLQNEDKIIISNPADS 240
QY 241 LIRTERPNNKEIVRYFIRHNALRAGTGENAPWVDEDELVKYSLPSKFSDFLLDPKYMT 300
Db 241 LIRTERPNNKEIVRYFIRHNALRAGTGENAPWVDEDELVKYSLPSKFSDFLLDPKYMT 300
QY 301 LNPSTKRNTGSPDRKPKSKTDSNLSPLNPKLWCHVHLKSLSGSPLKVNKSNK 360
Db 301 LNPSTKRNTGSPDRKPKSKTDSNLSPLNPKLWCHVHLKSLSGSPLKVNKSNK 360
QY 361 SPEEHLEEMKMWSPNKLHTNFH1PKGPPAKKPGKSHDKPLKAKGRSGKILNGOKSTGN 420
Db 361 SPEEHLEEMKMWSPNKLHTNFH1PKGPPAKKPGKSHDKPLKAKGRSGKILNGOKSTGN 420
QY 421 SKSPKGLTKPTKQKQMTLLDMAGTQKMTAPRNSGGTPTRTSKPHKHLPPAALHLIA 480
Db 421 SKSPKGLTKPTKQKQMTLLDMAGTQKMTAPRNSGGTPTRTSKPHKHLPPAALHLIA 480
QY 481 YYKENKDRKRSALSCV1SKTARLLSSDRARLPEELRSVLQKRYELLEHKRWASKE 540
Db 481 YYKENKDRKRSALSCV1SKTARLLSSDRARLPEELRSVLQKRYELLEHKRWASKE 540
QY 541 EQRKYLKKBELKKEKAKERREKEMLEBKQKRYEDQELTGKNLPAFLVDTPE 600
Db 541 EQRKYLKKBELKKEKAKERREKEMLEBKQKRYEDQELTGKNLPAFLVDTPE 600
QY 601 GLPNTLFGDVANVVEFLSCYGLLLPDAQYPTTAVSLMEALSADGGFLYLNRLVILLQ 660
Db 601 GLPNTLFGDVANVVEFLSCYGLLLPDAQYPTTAVSLMEALSADGGFLYLNRLVILLQ 660
QY 661 TLLQTLQDEIAEDYGELGKMLSE1PLTLHVSSELVRLCLRSDVQSESGSDTDNDKDS 720
Db 661 TLLQTLQDEIAEDYGELGKMLSE1PLTLHVSSELVRLCLRSDVQSESGSDTDNDKDS 720
QY 721 AAFEDNEVQDEFLEKLTSEFFELTSEEKLIQILTALCHRLMTYSVQDHMETRQMSAEL 780
Db 721 AAFEDNEVQDEFLEKLTSEFFELTSEEKLIQILTALCHRLMTYSVQDHMETRQMSAEL 780
QY 781 WKERLAVLKEENDKGAEKQKKEKEMAKNKENGKVENGLGKTDKRIKIVKPEPQVDTAE 840
Db 781 WKERLAVLKEENDKGAEKQKKEKEMAKNKENGKVENGLGKTDKRIKIVKPEPQVDTAE 840
QY 841 DMISAVKSRLLAIOAKERE1QEREMKVLQRQAEERI1RKHAAAKAFOEGIAKAKL 900
Db 841 DMISAVKSRLLAIOAKERE1QEREMKVLQRQAEERI1RKHAAAKAFOEGIAKAKL 900
QY 901 VMRTP1GTDNRNHNRYLWFLSDVPGLFIEKGWVHDSIDYRFNHHCDDHTVSGDEDCPRS 960
Db 901 VMRTP1GTDNRNHNRYLWFLSDVPGLFIEKGWVHDSIDYRFNHHCDDHTVSGDEDCPRS 960
QY 961 KXANLGKNSMNTQGTATEVAVETTPKQGNLWFLCDSQKELDELLNCLHPQIGRESQ 1020
Db 961 KXANLGKNSMNTQGTATEVAVETTPKQGNLWFLCDSQKELDELLNCLHPQIGRESQ 1020
QY 1021 LKERLEKRYQDI1IHS1HLARKENLGLKSCDGNQELLNFLRSPL1EVAITRLQKGGLYVEE 1080
Db 1021 LKERLEKRYQDI1IHS1HLARKENLGLKSCDGNQELLNFLRSPL1EVAITRLQKGGLYVEE 1080
QY 1081 TSEFEARV1SLEKLDGFCVIALQASV1KKFLQGFMAPKQKRRKLQSDSDSAKTEVDDEE 1140
Db 1081 TSEFEARV1SLEKLDGFCVIALQASV1KKFLQGFMAPKQKRRKLQSDSDSAKTEVDDEE 1140

QY 1141 KMVEEAKVASALEKWKTAIREAQTFSRMHVLLGLMDACIKWDMSAENARCKVCPKKGED 1200
Db 1141 KMVEEAKVASALEKWKTAIREAQTFSRMHVLLGLMDACIKWDMSAENARCKVCPKKGED 1200
QY 1201 DKLIJLDCENKAFHLCFLRPALEYVPGDQWOCQACQAPATARNRSGRNYTEESASEDSED 1260
Db 1201 DKLIJLDCENKAFHLCFLRPALEYVPGDQWOCQACQAPATARNRSGRNYTEESASEDSED 1260
QY 1261 DESDEEBEEDDEEDYEVAGLRPRKTIIRGHKSVIPPAARSGRRPGKPKHSTRSQ 1320
Db 1261 DESDEEBEEDDEEDYEVAGLRPRKTIIRGHKSVIPPAARSGRRPGKPKHSTRSQ 1320
QY 1321 PKAPPVDDAEDVDELVLOTKRSSRQSLQKCEELHKKI1VKYRSPWPPEPVTREADEBY 1380
Db 1321 PKAPPVDDAEDVDELVLOTKRSSRQSLQKCEELHKKI1VKYRSPWPPEPVTREADEBY 1380
QY 1381 YDVITHPMDFOTVQNKSCGYSYRSQVEFLTDKQVFTNAEVYNCRGSHVLSQWVTEOCL 1440
Db 1381 YDVITHPMDFOTVQNKSCGYSYRSQVEFLTDKQVFTNAEVYNCRGSHVLSQWVTEOCL 1440
QY 1441 VLLHKLHPGHYPYVRRKPKFPDLAEDGSEPEAVGQSDRDEDRRSREAEIQEWLQDTS 1500
Db 1441 VLLHKLHPGHYPYVRRKPKFPDLAEDGSEPEAVGQSDRDEDRRSREAEIQEWLQDTS 1500
QY 1501 LYSAKINSKDHNCFMMLVNTQFCWALTDTVT 1531
Db 1501 LYSAKINSKDHNCFMMLVNTQFCWALTDTVT 1531

RESULT 2

US-09-839-479-29
; Sequence 29, Application US/09839479
; Patent No. 6727222
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042002
; CURRENT APPLICATION NUMBER: US/09/839,479
; PRIOR FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 29
; LENGTH: 1531
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-839-479-29

Query Match 100.0%; Score 7985; DB 4; Length 1531;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1531; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAPLLGRKPPFLVNPPLGPEPFFTIPTHTQEAFTREEYEARELRYSERIWTCKTSSQOL 60
Db 1 MAPLLGRKPPFLVNPPLGPEPFFTIPTHTQEAFTREEYEARELRYSERIWTCKTSSQOL 60
QY 61 THKEAWEEOQVAELLKEEFPWYKLVLEVMVHNTASLEKLVDTAWLEIMTKYAVGSEC 120
Db 61 THKEAWEEOQVAELLKEEFPWYKLVLEVMVHNTASLEKLVDTAWLEIMTKYAVGSEC 120
QY 121 DFEVGKEMKLVKIKIHPLEKVDDEATEKSGDGCADSPSSDKENSSOIAQDHQKKTWV 180
Db 121 DFEVGKEMKLVKIKIHPLEKVDDEATEKSGDGCADSPSSDKENSSOIAQDHQKKTWV 180
QY 181 KEDEGRRESINDRARRSPKLTSLKGERKWAPPKFLPHKYDVKLQNEDKIIISNPADS 240

Db 181 KDEGRRESINDRARRSPKLPSTLSKGERKWAAPPKFLPHKVDVQLQNEDEK1ISNVPADS 240
QY 241 LIRTERPPNKEIIVRYEIRRNALRAGTGENAPWVEDELVKCYSLPSKFSDFLLDPYKMT 300
Db 241 LIRTERPPNKEIIVRYEIRRNALRAGTGENAPWVEDELVKCYSLPSKFSDFLLDPYKMT 300
QY 301 LNPSTRKNTGSDPRKPSKSKTDNSSLSSPLNPKLWCHVHLKKSLSGSPLVKXNSKNSK 360
Db 301 LNPSTRKNTGSDPRKPSKSKTDNSSLSSPLNPKLWCHVHLKKSLSGSPLVKXNSKNSK 360
QY 361 SPEEHLEEMKXNSPNKLTNTHFI PKKGPAPKPKGKSDKPLKAKGRSGIINGQKSTGN 420
Db 361 SPEEHLEEMKXNSPNKLTNTHFI PKKGPAPKPKGKSDKPLKAKGRSGIINGQKSTGN 420
QY 421 SKSPKGLTPTKTKMOMTLLDWAQTKQWTRAPRNSGGTPTRTSSKPHKHLPPAALHLIA 480
Db 421 SKSPKGLTPTKTKMOMTLLDWAQTKQWTRAPRNSGGTPTRTSSKPHKHLPPAALHLIA 480
QY 481 YYKENKDREDKRSALSCVISTARLLSSEDRARLPEELRSLVQKRYELLEHKRWASME 540
Db 481 YYKENKDREDKRSALSCVISTARLLSSEDRARLPEELRSLVQKRYELLEHKRWASME 540
QY 541 EORKEYLKKRBELKKLKEKAKERREKEMLEKQKRYEDELTKGNLPAFRLVDTPPE 600
Db 541 EORKEYLKKRBELKKLKEKAKERREKEMLEKQKRYEDELTKGNLPAFRLVDTPPE 600
QY 601 GLPNTLFGDVAWVVEFLSCYSGLLLPDAQYPTAVSLMEALSADKGGFLYLNARVLILQ 660
Db 601 GLPNTLFGDVAWVVEFLSCYSGLLLPDAQYPTAVSLMEALSADKGGFLYLNARVLILQ 660
QY 661 TLJLQTLQDEIAEDYELGELGWLSEIPLTLHSVSELVRLCLRRSDVQEESEGSDTDNKS 720
Db 661 TLJLQTLQDEIAEDYELGELGWLSEIPLTLHSVSELVRLCLRRSDVQEESEGSDTDNKS 720
QY 721 AAFEDNEVODEFLEKLETSSEFELTSEELKQILTALCHRLMTYSVQDHMETRQMSAEL 780
Db 721 AAFEDNEVODEFLEKLETSSEFELTSEELKQILTALCHRLMTYSVQDHMETRQMSAEL 780
QY 781 WKERLAVLKEENDKKAEBKQKMEAKNKENGKVGKTKDRKKRIVKFEPQVDTEAE 840
Db 781 WKERLAVLKEENDKKAEBKQKMEAKNKENGKVGKTKDRKKRIVKFEPQVDTEAE 840
QY 841 DMISAVKSRLLAIQAQKEREIQEREMKVLERQAEEREIRKHAAAEKAFQEGIAKAL 900
Db 841 DMISAVKSRLLAIQAQKEREIQEREMKVLERQAEEREIRKHAAAEKAFQEGIAKAL 900
QY 901 VMRRTPIGTDNRNRYWLFSDVPGLFIEKGWVHDSIDYRFNHHCKDHTVSGDEDYCPRS 960
Db 901 VMRRTPIGTDNRNRYWLFSDVPGLFIEKGWVHDSIDYRFNHHCKDHTVSGDEDYCPRS 960
QY 961 KKANLGKNASMTQHGTA TEVA VETTTPKOGNLWFLCDSQKELDELLNCLHPQGIRESQ 1020
Db 961 KKANLGKNASMTQHGTA TEVA VETTTPKOGNLWFLCDSQKELDELLNCLHPQGIRESQ 1020
QY 1021 LKERLEKRYQDIITHSLARKPNLGLKSCDGOELLNFLRSDLEI VATRLQKGLGYBEE 1080
Db 1021 LKERLEKRYQDIITHSLARKPNLGLKSCDGOELLNFLRSDLEI VATRLQKGLGYBEE 1080
QY 1081 TSEFEARVJLSEKLDKDFGECVIALQASVIKKFLQGFMAPKQRRKLOSDS AKTEVEUDE 1140
Db 1081 TSEFEARVJLSEKLDKDFGECVIALQASVIKKFLQGFMAPKQRRKLOSDS AKTEVEUDE 1140
QY 1141 KKVVEAKVASALEKWKTAI REAQTFSRMHVLGLMDACIKWDMSAENARCKVCPKGGED 1200
Db 1141 KKVVEAKVASALEKWKTAI REAQTFSRMHVLGLMDACIKWDMSAENARCKVCPKGGED 1200
QY 1201 DKLLCDECNKAFHFLCLRPALYEVDPGEWQCPACOPATARNRSGRNTYTESASDS 1260
Db 1201 DKLLCDECNKAFHFLCLRPALYEVDPGEWQCPACOPATARNRSGRNTYTESASDS 1260
QY 1261 DESDEEEEBEEDYEVAGLRPRKRTIRGKHSVIPPAAASGRRRFGKXPHSTRRSQ 1320

Db 1261 DESDEEEEBEEDYEVAGLRPRKRTIRGKHSVIPPAAASGRRRFGKXPHSTRRSQ 1320
QY 1321 PKAPPVDDAEVDDELVLQTKRSTRRSQLELQKCEEILHKIKYKRFSPFPREPVRDEADY 1380
Db 1321 PKAPPVDDAEVDDELVLQTKRSTRRSQLELQKCEEILHKIKYKRFSPFPREPVRDEADY 1380
QY 1381 YDVIITHPMDFTQVQNKSCGSYRSVQEFJLTMKQVFTNAEVYNCGRSHVLSQWVKTEQCL 1440
Db 1381 YDVIITHPMDFTQVQNKSCGSYRSVQEFJLTMKQVFTNAEVYNCGRSHVLSQWVKTEQCL 1440
QY 1441 VYLLHKHLPCHPYPVRKPKFPDLRAEDGDESEPAVCGOSRDEDRRSREABIOEWLODTS 1500
Db 1441 VYLLHKHLPCHPYPVRKPKFPDLRAEDGDESEPAVCGOSRDEDRRSREABIOEWLODTS 1500
QY 1501 LYSAKINSKDHNCFMMLVNTQFCMALTDTVT 1531
Db 1501 LYSAKINSKDHNCFMMLVNTQFCMALTDTVT 1531
RESULT 3
US-09-418-710-27
; Sequence 27, Application US/09418710
; Patent No. 6596482
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042001
; CURRENT APPLICATION NUMBER: US/09/418,710
; CURRENT FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 27
; LENGTH: 1527
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-418-710-27
Query Match 99.6%; Score 7955; DB 4; Length 1527;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 1527; Conservative 0; Mismatches 0; Indels 4; Gaps 1;
QY 1 MAPLGRKPPPLVNPDPGEEPPFTTPTQEAAPRTREEYEARLERYSERIWTCKSTGSSQL 60
Db 1 MAPLGRKPPPLVNPDPGEEPPFTTPTQEAAPRTREEYEARLERYSERIWTCKSTGSSQL 60
QY 61 THKEAWBEQEAELLKEEFPRAWYKLVLEWVHNTASLEKLVDTAWLEIMTKYAVGEEC 120
Db 61 THKEAWBEQEAELLKEEFPRAWYKLVLEWVHNTASLEKLVDTAWLEIMTKYAVGEEC 120
QY 121 DPEVGEKMWLVKVIKIHIPLEKVDSEATEKSDGACDSSDKENSSQIAQDHQKKEVTV 180
Db 121 DPEVGEKMWLVKVIKIHIPLEKVDSEATEKSDGACDSSDKENSSQIAQDHQKKEVTV 180
QY 181 KEDEGRRESINDRARRSPKLPSTLSKGERKWAAPPKFLPHKVDVQLQNEDEK1ISNVPADS 240
Db 181 KEDEGRRESINDRARRSPKLPSTLSKGERKWAAPPKFLPHKVDVQLQNEDEK1ISNVPADS 240
QY 241 LIRTERPPNKEIIVRYEIRRNALRAGTGENAPWVEDELVKCYSLPSKFSDFLLDPYKMT 300
Db 241 LIRTERPPNKEIIVRYEIRRNALRAGTGENAPWVEDELVKCYSLPSKFSDFLLDPYKMT 300
QY 301 LNPSTRKNTGSDPRKPSKSKTDNSSLSSPLNPKLWCHVHLKKSLSGSPLVKXNSKNSK 360
Db 301 LNPSTRKNTGSDPRKPSKSKTDNSSLSSPLNPKLWCHVHLKKSLSGSPLVKXNSKNSK 360
QY 361 SPEEHLEEMKXNSPNKLTNTHFI PKKGPAPKPKGKSDKPLKAKGRSGIINGQKSTGN 420

Db 361 SPEEHLLEEMKWMSPNKLHTNFHIPKGGPPAKPKGKSHDKPLKAGRSKGILNGKSTGN 420

Qy 421 SKSPKGLKTPKTKQKQWTLDMAGKTQKWTAPRNSGGTPTSTSKPHKHLPPAALHLIA 480

Db 421 SKSPKGLKTPKTKQKQWTLDMAGKTQKWTAPRNSGGTPTSTSKPHKHLPPAALHLIA 480

Qy 481 YYKENKORDEKFSALSVCVISTARTLLSSDRARLPEELRSLVQKRYELLEHKKRWASME 540

Db 481 YYKENKORDEKFSALSVCVISTARTLLSSDRARLPEELRSLVQKRYELLEHKKRWASME 540

Qy 541 EQRKEYLKKREBELKKLKEKAKERREKEMLEKQKRYEDELTKGNLPAFRLVDTPPE 600

Db 541 EQRKEYLKKREBELKKLKEKAKERREKEMLEKQKRYEDELTKGNLPAFRLVDTPPE 600

Qy 601 GLPNTLFGDVAMVBFSLCYSGILLIPDAQYPTAVSLMEALSADGKGFFLYLNRVLVILQ 660

Db 601 GLPNTLFGDVAMVBFSLCYSGILLIPDAQYPTAVSLMEALSADGKGFFLYLNRVLVILQ 660

Qy 661 TLLQTLLODEIAEDYGELGKMLSEIPLTLHVSSELVRLCLRSDVQEESEGSDDTDNKS 720

Db 658 -LLQTLLODEIAEDYGELGKMLSEIPLTLHVSSELVRLCLRSDVQEESEGSDDTDNKS 716

Qy 721 AAFEDNEVODEFLEKLETSEFFELTSEELQILITLALCHRLMTYSVQDHMETRQMSAEL 780

Db 717 AAFEDNEVODEFLEKLETSEFFELTSEELQILITLALCHRLMTYSVQDHMETRQMSAEL 776

Qy 781 WKERLAVLKEENDKKRAEKQKQKEMKAVKENGKVENGLGKTDRKKRIVKFPQVDTRAB 840

Db 777 WKERLAVLKEENDKKRAEKQKQKEMKAVKENGKVENGLGKTDRKKRIVKFPQVDTRAB 836

Qy 841 DMSAVKSRRLAIQAKKERETOEREMKVKLROAEEERIRKHAAAKAFOEGTAKAKL 900

Db 837 DMSAVKSRRLAIQAKKERETOEREMKVKLROAEEERIRKHAAAKAFOEGTAKAKL 896

Qy 901 VMRRTPIGTDNRHNYWLFSDVPGLFIEKGWVHDSIDYRFNHHCKDHTVSGDEDCPRS 960

Db 897 VMRRTPIGTDNRHNYWLFSDVPGLFIEKGWVHDSIDYRFNHHCKDHTVSGDEDCPRS 956

Qy 961 KXANLGKNSAMNTQGTATEVAVETTTPKQGNLWFLCDSQKELDELLNCLHPQIGRSQ 1020

Db 957 KXANLGKNSAMNTQGTATEVAVETTTPKQGNLWFLCDSQKELDELLNCLHPQIGRSQ 1016

Qy 1021 LKERLEKRYQDILHSTHLARKENLGLKSCDGNQELLNFLRSLIEVATRLQGGGLGYVEE 1080

Db 1017 LKERLEKRYQDILHSTHLARKENLGLKSCDGNQELLNFLRSLIEVATRLQGGGLGYVEE 1076

Qy 1081 TSEFEARVLSLEKLDGFCVIALQASVIKKFLQGFMAPKQRRKLQSDSAKTEBEVDDE 1140

Db 1077 TSEFEARVLSLEKLDGFCVIALQASVIKKFLQGFMAPKQRRKLQSDSAKTEBEVDDE 1136

Qy 1141 KQWTEAKVASALEKWKTAIRBAQTFSRMHVLLGLMDACIKWMSAENARCKVCPKKGED 1200

Db 1137 KQWTEAKVASALEKWKTAIRBAQTFSRMHVLLGLMDACIKWMSAENARCKVCPKKGED 1196

Qy 1201 DKLLILCDECNKAFHLCFLRPALYVDPDGEWQCPACOPATARNRSGRNYTSSASEDSED 1260

Db 1197 DKLLILCDECNKAFHLCFLRPALYVDPDGEWQCPACOPATARNRSGRNYTSSASEDSED 1256

Qy 1261 DESDEEEEBEEEEEDYEAVAGLRPRKTTIRGKHSVTPPAARSGRRPKGPHSTRSQ 1320

Db 1257 DESDEEEEBEEEEEDYEAVAGLRPRKTTIRGKHSVTPPAARSGRRPKGPHSTRSQ 1316

Qy 1321 PKAPPVDDAEVDDELVLQTKRSRRSOSLELOKCEEILHKTVKYPNSWFPREPVRDEADY 1380

Db 1317 PKAPPVDDAEVDDELVLQTKRSRRSOSLELOKCEEILHKTVKYPNSWFPREPVRDEADY 1376

Qy 1381 YDVIITHPMDFOVQNKSCGYSRVSQEFLLTDMKQVFTNAEVNCRGSHVLSQWVTEOCL 1440

Db 1377 YDVIITHPMDFOVQNKSCGYSRVSQEFLLTDMKQVFTNAEVNCRGSHVLSQWVTEOCL 1436

Qy 1441 VYLLHKHGLPHGPYVRKRKKFPDRDLAEDGSDSEPAVQSGRDEDRRSREAEIQEWLQDTS 1500

Db 1437 VYLLHKHGLPHGPYVRKRKKFPDRDLAEDGSDSEPAVQSGRDEDRRSREAEIQEWLQDTS 1496

Qy 1501 LYSAKINSKDHNCFMMLVNTQFCMALTDVT 1531

Db 1497 LYSAKINSKDHNCFMMLVNTQFCMALTDVT 1527

RESULT 4

US-09-839-479-27

; Sequence 27, Application US/09839479

; Patent No. 6727222

; GENERAL INFORMATION:

; APPLICANT: Jones, Michael H.

; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR

; FILE REFERENCE: 06501-042002

; CURRENT APPLICATION NUMBER: US/09/839,479

; CURRENT FILING DATE: 2001-04-20

; PRIOR APPLICATION NUMBER: US 09/418,710

; FILING DATE: 1999-10-15

; PRIOR APPLICATION NUMBER: PCT/JP98/01783

; PRIOR FILING DATE: 1998-04-17

; PRIOR APPLICATION NUMBER: JP 9/310027

; PRIOR FILING DATE: 1997-10-24

; PRIOR APPLICATION NUMBER: JP 9/116570

; PRIOR FILING DATE: 1997-04-18

; NUMBER OF SEQ ID NOS: 72

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 27

; LENGTH: 1527

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-839-479-27

Query Match 99.6%; Score 7955; DB 4; Length 1527;

Best Local Similarity 99.7%; Pred. No. 0;

Matches 1527; Conservative 0; Mismatches 0; Indels 4; Gaps 1;

Qy 1 MAPLIGRPPFLVNPPLPCEEPFTTIPHTQEAFTREEYEARLERYSERIWTCKSTGSSQL 60

Db 1 MAPLIGRPPFLVNPPLPCEEPFTTIPHTQEAFTREEYEARLERYSERIWTCKSTGSSQL 60

Qy 61 THKEAWBEEQVEAEILLKEEFPAYWEKLVLEMVHNTASLEKLVDTAWLEIMTKYAVGSEC 120

Db 61 THKEAWBEEQVEAEILLKEEFPAYWEKLVLEMVHNTASLEKLVDTAWLEIMTKYAVGSEC 120

Qy 121 DFEVGEKQKLVKVIHPILEKVDDEATEKSDGACDSPSSDKENSSQIAQDHQKQKTVV 180

Db 121 DFEVGEKQKLVKVIHPILEKVDDEATEKSDGACDSPSSDKENSSQIAQDHQKQKTVV 180

Qy 181 KEDEGRRESINDRARRSPKLPSTSLKGERKWAAPPKFLPHKYVDVKLQNEDKIISNPADS 240

Db 181 KEDEGRRESINDRARRSPKLPSTSLKGERKWAAPPKFLPHKYVDVKLQNEDKIISNPADS 240

Qy 241 LIRTERPPNKEIVRYFIRHNALRAGTGENAPWVVEDELVKVYSLPSKFSDFLLDPYKYMT 300

Db 241 LIRTERPPNKEIVRYFIRHNALRAGTGENAPWVVEDELVKVYSLPSKFSDFLLDPYKYMT 300

Qy 301 LNPSTKRNVTGSDPRKPSKSKTNDSSSLSPNPKLWCHVHLKKSLSGSPKVKNSKNSK 360

Db 301 LNPSTKRNVTGSDPRKPSKSKTNDSSSLSPNPKLWCHVHLKKSLSGSPKVKNSKNSK 360

Qy 361 SPEEHLLEEMKWMSPNKLHTNFHIPKGGPPAKPKGKSHDKPLKAGRSKGILNGKSTGN 420

Db 361 SPEEHLLEEMKWMSPNKLHTNFHIPKGGPPAKPKGKSHDKPLKAGRSKGILNGKSTGN 420

Qy 421 SKSPKGLKTPKTKQKQWTLDMAGKTQKWTAPRNSGGTPTSTSKPHKHLPPAALHLIA 480

Db 421 SKSPKGLKTPKTKQKQWTLDMAGKTQKWTAPRNSGGTPTSTSKPHKHLPPAALHLIA 480

Qy 481 YYKENKORDEKFSALSVCVISTARTLLSSDRARLPEELRSLVQKRYELLEHKKRWASME 540

Db 481 YYKENKORDEKFSALSVCVISTARTLLSSDRARLPEELRSLVQKRYELLEHKKRWASME 540

Qy 541 EQRKEYLKKREELKKLKEKAKERREKEMLEKQKRYEDELTKGNLPAFRLVDTPPE 600

Db 541 EQRKEYLKKREBELKKLKEKAKERREKEMLEKQRYDEQELTGKNLPAPFLVDTPTE 600
Qy 601 GLPNTLFGDVAMVVEFLSCYSGLLLPDAQYPTAVSLMEALSADKGGFLYLNRLVLLQ 660
Db 601 GLPNTLFGDVAMVVEFLSCYSGLLLPDAQYPTAVSLMEALSADKGGFLYLNRLVLI --- 657
Qy 661 TLLQTLQDEIAEDYBELGMLKSEIPLTLHVSSELVRLCLRRSDVQEESEGSDTDNDKDS 720
Db 658 -LLQTLQDEIAEDYBELGMLKSEIPLTLHVSSELVRLCLRRSDVQEESEGSDTDNDKDS 716
Qy 721 AAFEDNEVQDEFLEKLETSEFFELTSEKLIQILTALCHRLMTYSVQDHMETRQOMSABL 780
Db 717 AAFEDNEVQDEFLEKLETSEFFELTSEKLIQILTALCHRLMTYSVQDHMETRQOMSABL 776
Qy 781 WKERLAVLKEENDKKAQKQKREKEMAKNENKENGKVTGKTKRKRIIVKFEPOVDTEAR 840
Db 777 WKERLAVLKEENDKKAQKQKREKEMAKNENKENGKVTGKTKRKRIIVKFEPOVDTEAR 836
Qy 841 DMISAVKSRLLLAIOAKKEREIOEREMKVLERQAEERIRKHAAAFAFQEGIAKAL 900
Db 837 DMISAVKSRLLLAIOAKKEREIOEREMKVLERQAEERIRKHAAAFAFQEGIAKAL 896
Qy 901 VMERTDICTDRNHNRYLWFSDEVPGLFIEKGWVHDSIDYRFNHHCKDHTVSGDEDYCPRS 960
Db 897 VMERTDICTDRNHNRYLWFSDEVPGLFIEKGWVHDSIDYRFNHHCKDHTVSGDEDYCPRS 956
Qy 961 KXANLGNASMTQHGTAATEVA VETTT PKQGNLWFLCDSQKELDELLNCLHPQGRSQ 1020
Db 957 KXANLGNASMTQHGTAATEVA VETTT PKQGNLWFLCDSQKELDELLNCLHPQGRSQ 1016
Qy 1021 LKERLEKRYQDIHSHILARKPNLGLKSCDGNQELNFLRSDLIEVATRLQKGLGYVEE 1080
Db 1017 LKERLEKRYQDIHSHILARKPNLGLKSCDGNQELNFLRSDLIEVATRLQKGLGYVEE 1076
Qy 1081 TSEFEARVLSLEKLDQFGEVIALQASVIKKFLOGFWAPKQKRLQSDSASAKTEBVDSE 1140
Db 1077 TSEFEARVLSLEKLDQFGEVIALQASVIKKFLOGFWAPKQKRLQSDSASAKTEBVDSE 1136
Qy 1141 KQWVEAKVASALEKWKTAIREAQTFSRMHVLLGLMDACIKWDMSEANARCKVCPKKGED 1200
Db 1137 KQWVEAKVASALEKWKTAIREAQTFSRMHVLLGLMDACIKWDMSEANARCKVCPKKGED 1196
Qy 1201 DKLILCDECNKAFHLFCLRPALYEVDPGEWQCPACOPATARNRSRGNYTEBSASDS 1260
Db 1197 DKLILCDECNKAFHLFCLRPALYEVDPGEWQCPACOPATARNRSRGNYTEBSASDS 1256
Qy 1261 DESDEEBEEDYEVAGLRPRKTTIRGKHSVIPPAAARSRRPGKXPHSTRRSQ 1320
Db 1257 DESDEEBEEDYEVAGLRPRKTTIRGKHSVIPPAAARSRRPGKXPHSTRRSQ 1316
Qy 1321 PKAPPVDDAEVDLVLQTKRSRRQSLQKCEIILHKIVKRFSPRPPEVTRDAEDY 1380
Db 1317 PKAPPVDDAEVDLVLQTKRSRRQSLQKCEIILHKIVKRFSPRPPEVTRDAEDY 1376
Qy 1381 YDVITHPMDFTQVQNCSCGYSRSVQEFITDMKQVFTNAEVNCRGSHVLSQWVTEQCL 1440
Db 1377 YDVITHPMDFTQVQNCSCGYSRSVQEFITDMKQVFTNAEVNCRGSHVLSQWVTEQCL 1436
Qy 1441 VVLLHKHLPCHPYVRKRKFPDRLAEDGDSPEAVGQSRDRDRRSRABEIOEMLQDTS 1500
Db 1437 VVLLHKHLPCHPYVRKRKFPDRLAEDGDSPEAVGQSRDRDRRSRABEIOEMLQDTS 1496
Qy 1501 LYSAKINSKDHNCFMMLVNTQFCMALTDIVT 1531
Db 1497 LYSAKINSKDHNCFMMLVNTQFCMALTDIVT 1527

RESULT 5
US-09-418-710-69
; Sequence 69, Application US/09418710
; Patent No. 6596482
; GENERAL INFORMATION:

; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042001
; CURRENT APPLICATION NUMBER: US/09/418,710
; CURRENT FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 69
; LENGTH: 1525
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-418-710-69

Query Match 98.6%; Score 7876; DB 4; Length 1525;
Best Local Similarity 99.3%; Pred. No. 0;
Matches 1521; Conservative 1; Mismatches 2; Indels 8; Gaps 5;

Qy 1 MAPLLGRKPPPLVNPPLPGSEPPFTTIPHTQEAFTREYEARELYSERIWTCKTGSSQL 60
Db 1 MAPLLGRKPPPLVNPPLPGSEPPFTTIPHTQEAFTREYEARELYSERIWTCKTGSSQL 60
Qy 61 THKEAWEEOEVAELLKEEFPWYKGLVLEMHVHTASLEKLVDTAWLIMTKYAVGEEC 120
Db 61 THKEAWEEOEVAELLKEEFPWYKGLVLEMHVHTASLEKLVDTAWLIMTKYAVGEEC 120
Qy 121 DFEVGKEMLVKI VKIHPLEKVDSEATEKKGDCADSPSSDKENSSQIAQHQKKTVV 180
Db 121 DFEVGKEMLVKI VKIHPLEKVDSEATEKKGDCADSPSSDKENSSQIAQHQKKTVV 180
Qy 181 KDEGRRSINDRARRSPRKLPSTLKKGERKWAPKFLPHKYDVKLQNEDEKIIISNPADS 240
Db 181 KDEGRRSINDRARRSPRKLPSTLKKGERKWAPKFLPHKYDVKLQNEDEKIIISNPADS 240
Qy 241 LIRTERPNKELVYRPIRHNAIRAGTGENAPWVEDELVKYKSLSPKFSDFLLDPKYMT 300
Db 241 LIRTERPNKELVYRPIRHNAIRAGTGENAPWVEDELVKYKSLSPKFSDFLLDPKYMT 300
Qy 301 LNPSTKRNTGSPDRKPKSKXTDNSSLSPLNPKLWCHVHLKSLSGSPLKYVNSKNSK 360
Db 301 LNPSTKRNTGSPDRKPKSKXTDNSSLSPLNPKLWCHVHLKSLSGSPLKYVNSKNSK 360
Qy 361 SPEEHLEEMKQWSPNKLHNTNPHIPKGGPAKKPGKHSKPLKAKGRSKGILNGQKSTGN 420
Db 361 SPEEHLEEMKQWSPNKLHNTNPHIPKGGPAKKPGKHSKPLKAKGRSKGILNGQKSTGN 420
Qy 421 SKSPKKGLTKPTKMKQMTLLDMAGTQKMTAPRNSGGTPTSSKPHKGLPAPALHLIA 480
Db 421 SKSPKKGLTKPTKMKQMTLLDMAGTQKMTAPRNSGGTPTSSKPHKGLPAPALHLIA 480
Qy 481 YKKNKREDKRSALSCVISTARTLLSSEDRARLPEELRSLVQKRYELLEHKRWASMSSE 540
Db 481 YKKNKREDKRSALSCVISTARTLLSSEDRARLPEELRSLVQKRYELLEHKRWASMSSE 540
Qy 541 EQRKYLKKGREBELKKKKEKAKERREKEMLEBKQRYEDELTKGNLPAPFLVDTPTE 600
Db 541 EQRKYLKKGREBELKKKKEKAKERREKEMLEBKQRYEDELTKGNLPAPFLVDTPTE 600
Qy 601 GLPNTLFGDVAMVVEFLSCYSGLLLPDAQYPTAVSLMEALSADKGGFLYLNRLVLLQ 660
Db 601 GLPNTLFGDVAMVVEFLSCYSGLLLPDAQYPTAVSLMEALSADKGGFLYLNRLVLI --- 656
Qy 661 TLLQTLQDEIAEDYBELGMLKSEIPLTLHVSSELVRLCLRRSDVQEESEGSDTDNDKDS 720
Db 657 -LLQTLQDEIAEDYBELGMLKSEIPLTLHVSSELVRLCLRRSDVQEESEGSDTDNDKDS 715
Qy 721 AAFEDNEVQDEFLEKLETSEFFELTSEKLIQILTALCHRLMTYSVQDHMETRQOMSABL 780

Db 896 VMRTPIGDRNHNRYLWFSDEVPGLFIEKGWVHDSIDYRFNHCKDHTVSGEDYCPRS 955
Qy 961 KXANLGKNASMNTQHGTA TEAVETTTPKQGNLWFLCDSQKELDELLNCLHPQGIRESQ 1020
Db 956 KXANLGKNASMNTQHGTA TEAVETTTPKQGNLWFLCDSQKELDELLNCLHPQGIRESQ 1015
Qy 1021 LKERLEKRYQDIIHSHLARKPNLGLKSCDGNQELLNFLRSOLIEVATRLQKGLGYVEE 1080
Db 1016 LKERLEKRYQDIIHSHLARKPNLGLKSCDGNQELLNFLRSOLIEVATRLQKGLGYVEE 1075
Qy 1081 TSEFEARVLSLEKLDGFCVITALQASVKKFLOGFMAPKQKRKLQSDSAKTEEVDEE 1140
Db 1076 TSEFEARVLSLEKLDGFCVITALQASVKKFLOGFMAPKQKRKLQSDSAKTEEVDEE 1135
Qy 1141 KQWVEAKVASALEKWKTAIREAQTFSRMHVLLGMLDACIKWDMSAENARCKVCPKKGED 1200
Db 1136 KQWVEAKVASALEKWKTAIREAQTFSRMHVLLGMLDACIKWDMSAENARCKVCPKKGED 1195
Qy 1201 DXLILDECNKAFHFLCLRPALYVEPDGEWQCPACQATARNRSGRNYTEESASDSED 1260
Db 1196 DXLILDECNKAFHFLCLRPALYVEPDGEWQCPACQATARNRSGRNYTEESASDSEDSE- 1254
Qy 1261 DESDEEBEEREBEEDIEVAGLRPRKTRTGKHSVTPPAARSORRRPKKPHSTRRSQ 1320
Db 1255 DESDEEBEEREBEEDIEVAGLRPRKTRTGKHSVTPPAARSORRRPKKPHSTRRSQ 1314
Qy 1321 KPAPPVDDAEVDLVLQTKRSSRRQSLQKCEEIIHKIKVYRFSWPPREPVTREAEADY 1380
Db 1315 KPAPPV-DAEVDLVLQTKRSSRRQSLQKCEEIIHKIKVYRFSWPPREPVTREAEADY 1373
Qy 1381 YDVITHPMDQFQTVQNKCSGYSRSVQEFITDMKQVFTNAEVYNCRGSHVLSQWVKTEQCL 1440
Db 1374 YDVITHPMDQFQTVQNKCSGYSRSVQEFITDMKQVFTNAEVYNCRGSHVLSQWVKTEQCL 1433
Qy 1441 VVLLHKHLPCHPYVRKRKFPDRLAEDGDSPEAVGQSRDEDRRSREABIEQEWLQDTS 1500
Db 1434 VVLLHKHLPCHPYVRKRKFPDRLAEDGDSPEAVGQSRDEDRRSREABIEQEWLQDTS 1493
Qy 1501 LY-SAKINSKDHNCFMMLVNTQFCMALTDVT 1531
Db 1494 LYASAKINSKDHNCFMMLVNTQFCMALTDVT 1525

RESULT 7
US-09-949-016-7037
; Sequence 7037, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7037
; LENGTH: 1540
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-7037

Query Match 96.0%; Score 7662; DB 4; Length 1540;
Best Local Similarity 99.0%; Pred. No. 0;
Matches 1474; Conservative 2; Mismatches 7; Indels 6; Gaps 2;

Qy 1 MAPLLGRKPFPLVNPPLPGSEEPFTTIPHTQEAFTREERYEARLERYSERIWTCKTGSSQL 60
Db 58 MAPLLGRKPFPLVNPPLPGSEEPFTTIPHTQEAFTREERYEARLERYSERIWTCKTGSSQL 117
Qy 61 THKEAWEEBOEVAELLKEBFPAMYKGLVLEMHVHNTASLEKLVDTAWLMTKYAVGEEC 120
Db 118 THKEAWEEBOEVAELLKEBFPAMYKGLVLEMHVHNTASLEKLVDTAWLMTKYAVGEEC 177
Qy 121 DFEVGEKMLVKVIKVIHPLEKVDSEATEKSGDGACDSSPKDKENSSQIAQHQKKEVTV 180
Db 178 DFEVGEKMLVKVIKVIHPLEKVDSEATEKSGDGACDSSPKDKENSSQIAQHQKKEVTV 237
Qy 181 KDEGRRSINDRARRSPKLPSTLKKGERKWA PKFPHKYVDVQLQNEDKIISNVPADS 240
Db 238 KDEGRRSINDRARRSPKLPSTLKKGERKWA PKFPHKYVDVQLQNEDKIISNVPADS 297
Qy 241 LIRTERPPNKEIVRYFIRHNALRAGTGENAPWVEDELVKYKSLPSKFSDFLLDPYKYMT 300
Db 298 LIRTERPPNKEIVRYFIRHNALRAGTGENAPWVEDELVKYKSLPSKFSDFLLDPYKYMT 357
Qy 301 LNPSTKRKNTGSDPRKPSKSKTDNSSLSSPLNPKLWCHVHLKSLSGSPKLVYKNSKNSK 360
Db 358 LNPSTKRKNTGSDPRKPSKSKTDNSSLSSPLNPKLWCHVHLKSLSGSPKLVYKNSKNSK 417
Qy 361 SPEEHL EEMKMWSPNKLHTNFHIPKGPAPKPGKHSKDKPLAKAGRSKGI LNKQKSTGN 420
Db 418 SPEEHL EEMKMWSPNKLHTNFHIPKGPAPKPGKHSKDKPLAKAGRSKGI LNKQKSTGN 477
Qy 421 SKSPKGLKTPKTQKQMTLLDMAGTQKMTAPRNSGGTPTRTSSKPHKHLPAALHLIA 480
Db 478 SKSPKGLKTPKTQKQMTLLDMAGTQKMTAPRNSGGTPTRTSSKPHKHLPAALHLIA 537
Qy 481 YKKNKDRKRSALSCVISTARTLLSSEDRARLPEELRSLVQRYELLEHKRWASME 540
Db 538 YKKNKDRKRSALSCVISTARTLLSSEDRARLPEELRSLVQRYELLEHKRWASME 597
Qy 541 EQRKEYLKKRBEELKKLKEKAKERREKEMLEKQRYEQELTKGNLPAFLVDTPPE 600
Db 598 EQRKEYLKKRBEELKKLKEKAKERREKEMLEKQRYEQELTKGNLPAFLVDTPPE 657
Qy 601 GLPNTLFGDVAMVVEFLSCYSGLLIPDAQYPTAVSLMEALSADKGGFFLYLNRVLVILQ 660
Db 658 GLPNTLFGDVAMVVEFLSCYSGLLIPDAQYPTAVSLMEALSADKGGFFLYLNRVLVILQ 714
Qy 661 TLLQTLLODEIAEDYGEIGMKLSEIPLTLHVSVELVRLCLRRSDVQEESEGGSDTDNDKDS 720
Db 715 -LLQTLLODEIAEDYGEIGMKLSEIPLTLHVSVELVRLCLRRSDVQEESEGGSDTDNDKDS 773
Qy 721 AAFEDNEVQDEFLEKLETSEFFELTSEEKLIQILTALCHRLMTYSVQDHMETRQQMSAEL 780
Db 774 AAFEDNEVQDEFLEKLETSEFFELTSEEKLIQILTALCHRLMTYSVQDHMETRQQMSAEL 833
Qy 781 WKERLAVLKEENDKGRASKQKREKEMAKNKGKENGKVGKGLTKTRKRIVKFPQVDTEAE 840
Db 834 WKERLAVLKEENDKGRASKQKREKEMAKNKGKENGKVGKGLTKTRKRIVKFPQVDTEAE 893
Qy 841 DMI SAVKSRRLLAIOAKKEREIQEREMKVLQRQAEERIRKHAAAFAFQBGIAKAKL 900
Db 894 DMI SAVKSRRLLAIOAKKEREIQEREMKVLQRQAEERIRKHAAAFAFQBGIAKAKL 953
Qy 901 VMRRTPIGTDNRNHNRYLWFSDEVPGLFIEKGWVHDSIDYRFNHCKDHTVSGDEDYCPRS 960
Db 954 VMRRTPIGTDNRNHNRYLWFSDEVPGLFIEKGWVHDSIDYRFNHCKDHTVSGDEDYCPRS 1013
Qy 961 KXANLGKNASMNTQHGTA TEAVETTTPKQGNLWFLCDSQKELDELLNCLHPQGIRESQ 1020
Db 1014 KXANLGKNASMNTQHGTA TEAVETTTPKQGNLWFLCDSQKELDELLNCLHPQGIRESQ 1073
Qy 1021 LKERLEKRYQDIIHSHLARKPNLGLKSCDGNQELLNFLRSOLIEVATRLQKGLGYVEE 1080
Db 1074 LKERLEKRYQDIIHSHLARKPNLGLKSCDGNQELLNFLRSOLIEVATRLQKGLGYVEE 1133
Qy 1081 TSEFEARVLSLEKLDGFCVITALQASVKKFLOGFMAPKQKRKLQSDSAKTEEVDEE 1140

Db 1134 TSEFARVLSLEKLDKDFGECVIALQASVKKFLQFMQPKRRKLOSDSASKEVEDEE 1193
Qy 1141 KQWVEAKVASALEKWKTAIRAQAQTSFRMHVLLGMLDACIKWMSAENARCKVCPKKGED 1200
Db 1194 KQWVEAKVASALEKWKTAIRAQAQTSFRMHVLLGMLDACIKWMSAENARCKVCRKKGED 1253
Qy 1201 DKLLILDCENKAFHLFCLRPALYEVDPDGEWQCPACQPATARNRSGRNYTESASEDSED 1260
Db 1254 DKLLILDCENKAFHLFCLRPALYEVDPDGEWQCPACQPATARNRSGRNYTESASEDSED 1313
Qy 1261 DESDEEEEEEEDDEEYEVAGLRPRKTRGRKHSVTPPAARGRRPGKPPHSTRSQ 1320
Db 1314 DESDEEEEEEEDDEEYEVAGLRPRKTRGRKHSVTPPAARGRRPGKPPHSTRSQ 1373
Qy 1321 PKAPPVDDAEDVDELVTQKRSRRSRLQKCEIILHKIVKYRFSWPFREPVTRDEADY 1380
Db 1374 PKAPPVDDAEDVDELVTQKRSRRSRLQKCEIILHKIVKYRFSWPFREPVTRDEADY 1433
Qy 1381 YDVITHPMDFTQVQKSCGYSRSVQEFITDMKQVFTNAEVYVNCGRSHVLSWVKTQCL 1440
Db 1434 YDVITHPMDFTQVQKSCGYSRSVQEFITDMKQVFTNAEVYVNCGRSHVLSWVKTQCL 1493
Qy 1441 VVLLHKHLPCHPYVRKRRKFFDRLAEDGDESEPAVGOSRDEDRRSRE 1489
Db 1494 VVLLHKHLPCHPYVRKRRKFFDRLAEDGDESEPAVGOSR--GRRQK 1540

RESULT 8
US-09-418-710-1
; Sequence 1, Application US/09418710
; Patent No. 6536482
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042001
; CURRENT APPLICATION NUMBER: US/09/418, 710
; CURRENT FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1674
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-418-710-1

Query Match 11.7%; Score 930.5; DB 4; Length 1674;
Best Local Similarity 23.2%; Pred. No. 2.6e-58;
Matches 397; Conservative 263; Mismatches 532; Indels 521; Gaps 67;
Qy 3 PLLGRKPF----PLVNPDPGEPFPTIHTQAPFTRBEEYARLERYERIWTCKSTGSS 58
Db 152 PLLHRKPFVRQPPADLRPDDEEVFY-CVNTNIEFRHYDDFFERTILCNLSVMSCAVTRP 210
Qy 59 QLTHKEAMEEEOEVALLEKEEFPANVEKLVLEMVH-----HNTASLEKLVDTAWLEIMTKY 114
Db 211 GUTYQEALESSEKAKQNL--QSPF---EPLIIPVLVLTSLTHRSRLHEICDDIDFVAVVKORY 266
Qy 115 AVGECDCEVFGKMKLVKVIKIHPLEKVDDEATEKKGSDGCDSPSSDKENSSQIAQDHQ 174
Db 267 FVEETVEVIRNNGARLQCTILEVP-----PS-----HQNGFANGHV 303
Qy 175 KK---ETVVKDEGRRESINDARRSPKPLTSLKGGKRWAPPKPLPHKYDV----KLQ 227
Db 304 NSVDGTEIISDDSETQS-----CSFGNGKKKDAIDPLL-FKYKVQPTKKEL 351
Qy 228 NEDKIISNPADSLRTERPPNKEIVRYPIRNALRAGTGENAPWVVEDELVKYSLPS- 286

Db 352 HESAI---VKATQTSRRKHLSRDKLFLKQHC-----EPQEGVIK---IKASSLSTY 399
Qy 287 ----KFSDFLLDPYKMYTLNPSYTKRNTGSPDRKPSKK---SKTDSNLSPLNPKLWC 338
Db 400 KIAEQDFSYFFDDPPPTIFSPANRRRG-----RPPKRIHISQEDN----- 440
Qy 339 HVHLKXLSGSPKLVKXNSKNSKSPPEEHLEEMKXMSPNKLHTNPHI PKKGPAPKPKGHS 398
Db 441 -VANKQTLLA-----SYRSKATKR----- 458
Qy 399 DKPLKAKGRSGIILNGQSTGNSKSPKGLKTPKTKMKQMTLLDMAKGTQKMTAPRNSG 458
Db 459 DKLLK-----QEMKSLAFKAKLKR----- 479
Qy 459 GTPRTSSKPHKLPAAALHLIAYYKENKDRBKXSALSCVISTARTLSSSDRARKLPBEL 518
Db 480 -----EKADALEAKKKEKEDKEKR-----BEL 502
Qy 519 RSLVQKRYELLEHKKQWASMSSEQKEYLKKREELKKLKEKAKERREKEMLEKOK 578
Db 503 KKIVEEE-----RLKKKEKERLKVEREKEREKURE--EKRYVEYVKQWSKPR 549
Qy 579 RYEDQELTG-XNLPAFLRLVDTPEGLPNTLFGDVAMVVEFLSCYSGLLLPDAQP--ITAV 635
Db 550 --EDMECDLDELPEPTPKT--RLPPEIFGDALXVLEFLNAFGELFDLQDEFPDGVTL 605
Qy 636 SLMEAL-SADKGGFLYLNRLVILLQTLQLL--QDEIAE-----DYGELGKKSEIPLT 688
Db 606 VLEBALVNDSEG--PLCELLFFLTAIFAIEEVEEVEAKEQLTADTKGCSLXSLDLD 663
Qy 689 LHSVSELVRLCLRRSDVOEESGSDTD-----DNKDSAAFEDNEVODEFLEKLETSEF 741
Db 664 SCTLSEILRLHLASGADVTSANAKYRYQKGGFDATDDACMELSLNSPLVKKLSUSTSV 723
Qy 742 FELTSEEKLOILTALCHRLMTYSVQD-----HMETRQQMSA 778
Db 724 YDLTFGEKMKIILHALCGKLLTLVSTRDFIEDYVDILROAKQEFRELKAEQHRKEEAAA 783
Qy 779 ELWKERLAVLKEENDKGAEKOKKEMEAKN-----KENGKVENGLGTDRKKRIV 829
Db 784 RIRKKEKLEKEQEKMEKQEKLEKDEQRNSTADISIGEEREEDFTSIESKDEQK-- 841
Qy 830 KFEQVDTAEADMISAVKSRLLAIQAK-----KEREIOE-----REMVKLERQA-E 876
Db 842 ELDDQMFTEDEDDPGSHKRG-----RGKRGONGKFEFTRQEQINCVRTELTADEEEALK 897
Qy 877 EERIRKHAATAAKAFQEGIAKAKLVMMRTPIGTDNRNHRWYLFSDSEVPGLFTEK---GWV 933
Db 898 QEHQKEKELLEK-IQSAIACNTIP-----PLGRDRMYRYWIF-PSIFGLFIEDYSGLT 951
Qy 934 HDSI-----DYRFNHHCKDHTVSGDEDYCPKSKKAN-LGKNASMTQGTGATEVAVETT 987
Db 952 EDMLLPSPSSFQNNVQSDPQVS-----TKTGEPLMSBSTSNDQO-PRDHSVOLPK 1002
Qy 988 PQQONLWFLCDSQKELDELANCLHPQIGIRESOLKERL--EK-----RVQDIHSH 1037
Db 1003 PVHKPNRCFYSSCEQLDOLIEALNSRGRHSALKETLLOEKSRICQAQLARFSE--EKPH 1060
Qy 1038 LARKPNLGLK-----SCDGNQ-----ELNFLRSDLIEVATRIQKGLGYVEET- 1081
Db 1061 FSDKQPPSKTSYGRSSSNAYDPSQCAEQLELRDLRDLFLDIEDRIYQGLGAIKVD 1120
Qy 1082 -----SEFEA---RVISLEKL-----KDFGECVIALQASVTKKPLQGMAPKQKR 1124
Db 1121 RHWRSALESCKRYELLSBENKENGIIKTWNEDVEEMEIDEQTKVIVK--DRLLGIKTETP 1178
Qy 1125 KLOSDSASKEVEDE-----EKQWVEAKVAS-----ALEKWKTAI 1160
Db 1179 STVNASTPOSVSVSVHYLAWALFOIEQGIERRFLKAPLDASDSGRSVKTVLDWRBSL 1238
Qy 1161 REAOTFSRMHVLLGLMLDACIKWMSAENARCKVCPKKGEDDKLILCDBCNKAFHLFCULRP 1220


```
Db 1239 LSSASLSQVFLHJLSTLDRSVINSGKSLNARCKICRKKGDAENNVLCDCGCDRGHTTVCYVRP 1298
Qy 1221 ALYEPDGMQCPACOPATARRNSRGNTYTESASEDSEDSDEEE-----EEEEEE 1276
Db 1299 KLKTVPEGDFCEPKQRCRLSRQRPSELSDEDEVEDMGDEVDGDEEGQSEE 1358
Qy 1277 EDEYV-----AGLRUPRXTI-----ROKHSVIPPAARSRRPKKPHSTRR 1318
Db 1359 EYEVEDQEDDSQEEBEVSLPKRGPQVRLPVKTRGLSSSFSSRQCOQEPGRYPSRQ 1418
Qy 1319 SOPK-----APPVD-----DAEYDEL----- 1334
Db 1419 STPKTIVSSKTRSLRKINSAPTETKSLRIASRSTRHSHGPLQADVFVLLSPRRKRG 1478
Qy 1335 -----VLQKRSRRQSL----- 1347
Db 1479 RKSANNTPENPNFRVIATKSSQSRVSIASKLSLQESKRCRKRQSPSPVPT 1538
Qy 1348 -----ELQKCEILHKIVKYPSPFPREPVTTRDAEDYDVITHPMDPTVQ 1394
Db 1539 LGRSSGROGGVHELSAFEQLVVELVRHDDSWPFLKLSKIQVPDYDIKKPIALNIIR 1598
Qy 1395 NKCSGYSRSVOEFLDMKOVFTNAEYVNCRS 1427
Db 1599 EKVNCKEYKLASEFIDDIELMFSNCFEYINPRT 1631

RESULT 9
US-09-839-479-1
; Sequence 1, Application US/09839479
; Patent No. 672722
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042002
; CURRENT APPLICATION NUMBER: US/09/839,479
; PRIOR FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1674
; TYPE: PRN
; ORGANISM: Homo sapiens
US-09-839-479-1

Query Match 11.7%; Score 930.5; DB 4; Length 1674;
Best Local similarity 23.2%; Pred. No. 2.6e-58;
Matches 397; Conservative 263; Mismatches 532; Indels 521; Gaps 67;

Qy 3 PLLGRKPF-----PLVNPLGPERPFTTIPHTQAFRTREYEARLERYSIWTKSGSS 58
Db 152 PLLHRKPFVQRPPADLRDEEVFY-CKVTNEIFRHYDDFFERTILCNLSLVWSCAVTGRP 210
Qy 59 QLTKEAMEEEQVABLLKEEFPWYKLVLEMVH-----HNTASLEKLVDTAWLEIMTKY 114
Db 211 GLTYQEALESSEKARQNL-QSFP--EPLIIPVLVITSLTHSRULHEICDDIFAYVKORY 266
Qy 115 AVGECDFEVGKMKLVKVIKHPLEKVEDEATEKKSOGACDSPESSDKENSQIAQDHQ 174
Db 267 FVEETVEIRNNGARLQCTILEVLP-----PS-----HQNGFANGHV 303
Qy 175 KK---ETVYKDEGRRESINDRARSPRKLPSTLKKGERKWAPPKFLPHKYDV-----KLQ 227
Db 304 NSVDGETIILSDSDSETQS-----CSFQNGKKDAIDPLL-FKYKVQPTKCEL 351

Qy 228 NEDKIISNPADSLIRTPPNKEIVRYRIRNARAGTGENAPWVVEDELVKVYSLPS- 286
Db 352 HESAI---VKATOIGRRKHLPSRDKLKLKQHC-----EPQEGVIK---IKASSILSTY 399
Qy 287 ---KFSDFLLDPKYMTLNPSTKRNKGTSPDRPSKK---SKTDSNSSLSPNFKLWC 338
Db 400 KIAEQDFSYFFDDPPTFFISPANRRG-----RPPKRIHISQEDN----- 440
Qy 339 VHYLKKSLSGSLPKVKNKSNKSPBEHLBEEMKMMSPNKLHTNFHFKPKGPPAKPKGKHS 398
Db 441 -VANQTLA-----SYRSKATKER----- 458
Qy 399 DKPLKAKGRSKGILNGQSTGNSKPKGLKTPKTKMKQMTLLDMAKGTQKMTAPRNSG 458
Db 459 DKLLK-----QBEKMSLAPEKAKLKR----- 479
Qy 459 GTPRTSSKPHKLPPAALHLIAYYKENKDREKRSALS CVISKTARLLSSEDRARLPEEL 518
Db 480 -----EKADALEAKKKEKDEKKR-----EEL 502
Qy 519 RSLVQRYVELLEHKKRWASMSSEORKEYLKKREELKKLKEKAKERREKEMLERLEKOK 578
Db 503 KKIVEE-----RLKKEKERLKVEREKEREKLR--EKRKYVEVLKQMSKRP 549
Qy 579 RYEDQELTG-KNLPAFLVDTPEGLPNTLFGDVAMVVEFLSCYSGLLLPDAQYP--ITAV 635
Db 550 --EDMECDOLKELPEPTPKT--RLPPEIFGDALVLEFLNAPGELFDLQDFPDGVTLE 605
Qy 636 SLMEAL-SADKGGFLYLNARVILLQTLQTL--QDEIAE-----DYGLGHLKLSIPLT 688
Db 606 VLEEALVGNDSG--PLCELLFFFLTAIFQAIAEBEEVAKBQLTDADTKGCSLSLDLD 663
Qy 689 LHSVSELVRLCLRRSDVOEESGSDTD-----DNKDSAAFPEDNEVQDEFLEKLETSBF 741
Db 664 SCTLSEILRLHLASGADVTSANAKRYQKRGFPATDDACMELRLSNPSLVKLSSTSV 723
Qy 742 FELTSEEKLQILTALCHRLMTYSVOD-----HMETROQMSA 778
Db 724 YDLTPCKMKILHALCGKLLTLVSTRDFTIEDYVDILROAKQBFRELKAEQHRKEREAAA 783
Qy 779 ELWKERLAVLKEENDKRAEKOKRAMEAKN-----KENGKVENGLGKTDKRRIV 829
Db 784 RIRKKEEKLKEQEQKMEKQEKJEDQORNSTADISIGEEERDFDTSIESKDTQK-- 841
Qy 830 KPEPOVDTEADMISAVKSRRLAIOAK-----KEREIQE-----REMKVKLERQA-E 876
Db 842 ELDDQMTFDEDDPGSHKRG--RGKKGQNGFKEFTQEQINCVCVTRRELLTADEEALK 897
Qy 877 EERIRKHAABKAFQEGIAKAKLYMRRTPICTDRNHNRYWLFSDVPGLFIEK---GWV 933
Db 898 QEHQKKEKELLEK-IQSAIACNIF---PLGRDRMYRYWIF-PSIPGLFIEEDYSGLT 951
Qy 934 HDSI-----DYRPNHCKDHTVSGDEDYCPRSKKN-LGKNASMTQHTGATEVAVETT 987
Db 952 EDMLPRPSSFQNNVQSDQPVQS-----TKTGEPLMSESTSNIDQG-PRDHSVOLPK 1002
Qy 988 PKQGNLWFLCDSOKDELINCLHPQGIRESOLKRL--EK-----RYODITHSH 1037
Db 1003 PVHKPNRCFVSSCEQLQOLIEALNSRGHRESALAKETLLOEKSRICAQLARFSE--EKFH 1060
Qy 1038 LARKPNLGLK-----SCDGNQ-----ELNLPLRSDLIEVATRLQKGLGVYEET- 1081
Db 1061 FSDKFPQPSKPTYSRGRSSNAYDPSQCAEKQLELRDLDFLDIEDRIYQGTGAIKAVTD 1120
Qy 1082 -----SEFEA---RVISLEKL-----KDFGCVIALQASVKKFLOGFWAPKQKR 1124
Db 1121 RHIMRSALSGRYELLSEENKENGIIKTVDNEVEIMEIDQTKVIVK--DRLLGIKTETP 1178
Qy 1125 KLQSDSDAKTEERVDE-----EKKMVEEAKVAS-----ALEKWKTAI 1160
Db 1179 STVNSTASTQSVSSVHYVHVALMALFOIQEGIERRFLKAPLADSDSGRSYKTVLDWRRESL 1238
Qy 1161 REAQTFSRMHVLGLMDACIKWDMNAENARCKVCPKKGEDDKL1LCDECNKAFHLFLCLRP 1220
```


Db 1236 SLLSSASLSQVFLHLSTLDRSVIWSKSIILNARCKI CRKKGDAENMVLCDCGRGHHTYCV 1295
Qy 1219 RPALYVDPGEWQCPACQATARNRGRNRYTESASEDSEDDSEDEEEBEE 1274
Db 1296 RPKLKI VPEGDFCPCRPQRCRRLSFRQPSLESDDEDVDSMGDEDDVDSGEGOS 1355
Qy 1275 EEDYEV-----AGLRPRKTY-----RGKHSVIPPAAIRGRPRGKPKPHST 1316
Db 1356 EEEYEVQEDDSDSEEEVSLFKRGPQVRLPVKTRGKLSFSSRGQQQEPGRYPSPS 1415
Qy 1317 RRSOPK-----APPVD-----DAEVEL----- 1334
Db 1416 QOSTPKTVSSKTGRSLRKINSAPPTETKSLRTASRSTRSHSGPLOADVPFVLLSPRRKR 1475
Qy 1335 -----VLQTKSSRRQS----- 1347
Db 1476 RGRKSANNTPENSPNFRVIATKSSQSRSVNIASKLSQSESKRCKRQSPSP 1535
Qy 1348 -----ELQCEELHKIVKYRFSWPFPVTRDEADYDVITHPMDFOT 1392
Db 1536 VTLGRSSRGQGVHLSAFEQLVVELVRHDDSWPFLKLVSKIQVPDYDIIKKPIALNI 1595
Qy 1393 VQNKSCGVSRSQEFELTMKQVFTNAEVYVNCGRS 1427
Db 1596 IREKVNKCEYKLASEFIDDIELMFSCFETNPRT 1630

RESULT 11
US-09-839-479-69
; Sequence 69, Application US/09839479
; Patent No. 6727222
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042002
; CURRENT APPLICATION NUMBER: US/09/839,479
; CURRENT FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 69
; LENGTH: 1673
; TYPE: PRN
; ORGANISM: Homo sapiens
US-09-839-479-69

Query Match 11.5%; Score 919; DB 4; Length 1673;
Best Local Similarity 22.9%; Pred. No. 1.8e-57;
Matches 393; Conservative 262; Mismatches 534; Indels 526; Gaps 64;

Qy 3 PLLGRKPF-----PLVNPLCEBFFTPHTQEAFTREBYEARLERYSERIWTCKSTGSS 58
Db 152 PLLHRPFVQRPADLRDEEVFY-CKVTNEIFRHYDFFERTILCNLSLWMSCAVTGRP 210
Qy 59 QLTKEAWBEEQVALLKEEFPAYWEKLVLEMHV-----HNTASLEKLVDTAWLEIMTKY 114
Db 211 GLTYQEALESSEKARQNL-QSEF---EPLIIPVLYLTSLTHRSRLHEICDDIFAYVKDXY 266
Qy 115 AVGEBCDFVGEKMKLVKIVKIHLEKVDDEATEKKSQACDPSDKENSQIAQDHO 174
Db 267 FVEETVEIRNNGARLQCTILEVLP-----PS-----HQNGFANGHV 303
Qy 175 KK---ETVVKEDGRRESINDRARRSPRLPTSLKKGERKWAAPPKPLPHKYDV-----KLQ 227
Db 304 NSVDGETIITISDDSDSETQS-----CSFQNGKKDAIDPLL-FKYKQVPTKBL 351

Qy 228 NEDKIISNPADSLIRTERPPNKEIVRYFIRNNAAGTGENAPWVVBELVKYSLPS- 286
Db 352 HESAI---VKATQISRRKHLFSRDKLFLKQHC-----EPQGVIK---IKASSLSTY 399
Qy 287 -----KFSDFLLDPYKYMTLNPSTKRNKNGTSPPDRKPSK---SKTNSLSLSSPLNPKLWC 338
Db 400 KTAEQDFSYFFDDPTFFIFSPANRRG-----RPPKRIHISQEDN----- 440
Qy 339 HVHLKKSLSGSLPKVKNKSNKSPBEHLEEMMKMSPNKLHNFHFKPKGPPAKPKGKHS 398
Db 441 -VANKOTLA-----SYRSKATKER----- 458
Qy 399 DPLPLKAGRSKILNGQKSTGNSKPKGLKTPKTQKQMTLLDMAKGQKQMTAPRNSG 458
Db 459 DKLLK-----QEBMKSALFAKAKLKR----- 479
Qy 459 GTPRTSSKPKHLPPAALHLIAVYKENKDREKRSALSVCISKTAARLLSSEDPARLPEEL 518
Db 480 -----EKADALEAKKKEKDEKKR-----EEL 502
Qy 519 RSLVQKRYELLEHHKRWASMBEQRYLKKREELKKLKEKAKERREKEMLEKQK 578
Db 503 KKIVEE-----PLKKEKELKVEREKERLREE-----KKYVEYQWSK 546
Qy 579 RYEDQELTG-KNLPAFRLVDTPEGLPNTLFGDVAMVVEFLSCYSGLLLPDAQVP--ITAV 635
Db 547 PREDMECODLKELPEPTPVKT--RLPPEIFGDALVLEFLNAPGELFDLQDEPPDGVILE 604
Qy 636 SLMEAL-SADKGGFLYNRVILLQTLQTL--QDEIAE-----DYBELGMKLSIPLT 688
Db 605 VLEELVGNDSG--PLCELLFFFLTAIFQAIABEEVAKQOLTADTKGCSLKSLLDD 662
Qy 689 LHSVSELVRLCLRRSDVQSESGSDTD-----DNKDSAAFPEDNEVQDEFLEKLTSEF 741
Db 663 SCTLSEILRLHLIAGADVTSANAKYQKRGFPDATDACHMELRLSNLSLVKLSSTSV 722
Qy 742 FELTSEELQILTALCHRLMTYSVOD-----HMETROOMSA 778
Db 723 YDLTPGKMKILHALCGKLLTLVSTRDFIEDYVDILRQAKQBFRELKAEQHRKEREAAA 782
Qy 779 ELWKERLAVLKEENDKKAQKRAKQKMEAKN-----KENGKVENGLGKTDKRRIV 829
Db 783 RIRKKEEKLKEQEQKMEKQKLEKEDQORNSTADISIGEEREDPDTIESKDTQK-- 840
Qy 830 KFEPOVDTEAEDMISAVKSRRLAIAQK-----KERETQE-----REMKVKLERQA-E 876
Db 841 ELDQDMFTEDEDDPGSHKGR---RGKRGQNGFETFEQEQINCVCYTRLLTADAEEALK 896
Qy 877 EERIRKHAABAFQEGIAKAKLVNRRTPIGTDNRHNNRYWLFSDVPGFLTEK---GWV 933
Db 897 QEHQKKEKELLEK-LQSAIACTNIF---PLGRDRMYRRYWI-F-PSLPLGFTI EEDYSGLT 950
Qy 934 HDSI-----DYRFNHHCKDHTVSGDEDYCPRSKAN-LQKNASMTQHTGATEVAVETTT 987
Db 951 EDHLPRFPSSFQNNYQSDOPQS-----TKTGEPLMSESTSNIDQG-PRDHSVOLPK 1001
Qy 988 PRQGNLWFLCDSOKELDELLNCLHPQGIRESOLKRL----- 1025
Db 1002 PVHKRNWCFYSCQOLDQIIEALNSRGHRESALHETLLOEKSRIQAQLARSEEPKPS 1061
Qy 1026 EKRYQDIHSHILARKPNLGLKSCDGNQ-----ELNLFRLSDLIEVATRLQKGLGYVEE 1080
Db 1062 DKRPDSKFTYSGRSSN---AYDPSQMAEQLELRDLRDLFLDIEDRIYQGTGLAIKV 1117
Qy 1081 T-----SEFEA---RVISLEKL-----KDPGECVIALQASVKKFLOGFMAPKQK 1122
Db 1118 TDRIHWRSALESGRYELLSEENKENGIIKTVNEDEVEMEIDEQTKVIVK--DRLLGIKTE 1175
Qy 1123 RRLQSEDSAKTEEYDE-----EKKWVEEAKVAS-----ALEKWK 1158
Db 1176 TSTSTVSTNASTQSSVSVVHYLAMALFQIEQGLERRFLKAPLSDASGSRYSKTVLDRWE 1235
Qy 1159 AIREAQTFSRMHVLLGLMDACIKWDMNAENARCKVCPKKGEDDKLJLDCNKAFLFCL 1218

Db 770 EAKKKKEE-----AANAKLEAEKRIK-EKEMR---RQAVLLKHQER----- 809
Qy 540 EQRKEYLKKRBEKKLEKAKERREKEMLEKQKRYEAOEL---TGKNL----- 590
Db 810 ERRQHMWLMKAMEAKKAEERLKQEKDEKRLNKEKRLQORRLEWAKELKXPNED 869
Qy 591 -----PAFRLVDTPB-GLPNTLFGDMVAVVEFLSCYSGLLLPDQAQYPIITAVSLMEALS 642
Db 870 MCLADQKPLFELPRIFGLVLSGTSFSDCLMVQFLNFQKVLGDFVDVNPVLSVLQE-- 927
Qy 643 ADKGGFL---YLNRLVILLOTLLOLQDEAEDY---GELGMKLSIPLTLHVSSEL 695
Db 928 ----GLNLTGDSMGEVDQLLRLLSAVCDPGITGYKAKTALGEHLLNVGNRDNVSEI 983
Qy 696 VRLCLRRSDVQEESEGSDDNNKDSAAFEDNEVDQEFLEKLETSSEFFELTSBEKQILTA 755
Db 984 LQIFM-----EAHCQOTELTESLTKAKAQAHTPAQKASVLAF 1020
Qy 756 LCHRIILMTYSVDHMETROQMSAELWKLAVLKEENDKKRAEKQKQKEMEAQKENGKV 815
Db 1021 LINELACSKSVSEIDKNIDYMSNLRDRKVV---EGKRLKRLIIHAKTKGRDTSGGI 1076
Qy 816 ENG-----LG-----KTRKKRIVKFEPOVDTEAEDMISAVKSRLLALQAKKEREIOE 864
Db 1077 DLGEEQHPGLTTPPKRRRRKGGSDYDDDDDDSD-----QGDDEDEE 1123
Qy 865 --REMKVKLERQAEERIRKHAAREK-----AFOGIKAKLVNRRTPIGT 909
Db 1124 DREDQKGTDCEDDEGDQASVVELEKQIEKLSQOSQYRKLFDASHLSRVMFOP 1183
Qy 910 DRNHNRYWLPSEVPGLFIE-----KGWVH-----DSIDY 939
Db 1184 DRYRRRYWIL-PRCGIFVEGMSGEGLEIAKEREKKAERSVQIKEMFETSGDSLNC 1242
Qy 940 RFNHC-----KDHT---VSGDEDYCPRSK-----KANLGRNA----- 969
Db 1243 SNTDHCQEKEDLKBKNDTNLFLQKPGSFKLSKLEVAKMPPESEVMTPKPNAGANGCTL 1302
Qy 970 ----SMNTQHGTAATEVAE-----TTTPKQOQ 992
Db 1303 SYONGSKHSLGVSQSTATQSNVEKADSNLFWNTSGSGPKGYFSPPLNDQLLTKLEKRNQ 1362
Qy 993 NWFL-----CD----- 999
Db 1363 --WFSLLPRTPCDDTSLTHADMSTASLVTQSQPPSKSPSPTPAPLGSQAQNPVGLNPPA 1420
Qy 1000 ----- 999
Db 1421 LSPLOVKGVSWMGLQFCGWPTGVTNSIPTLSVPSLGLSGLSEGNNSFLTNSVASS 1480
Qy 1000 -----SOKELDELLNCLHP 1013
Db 1481 KSESPVPQNEKATSAQAAVEAKVPDPSPKPIPEMQFGHWRIIDPEDLKALLKVLHL 1540
Qy 1014 QGIRSQLKERLEKYQDIIHSIHLARKPNLGLKCDGNOELNLFRLSLDIE---VATRL 1070
Db 1541 RGIREKALQKQIK-HLDVITQACLKAK-DVAIIELNEE--NQVTRDIVENWSVEQA 1596
Qy 1071 QKGGLYVEETSEFEARVLSLEKLDKDFGECVIALQASVTKKFLQGMARQ----- 1121
Db 1597 MEMDLSVLQOQVEDLERRVAS-----ASLQ-----VKGWMCPEASEREDLVY 1638
Qy 1122 -----KRRKLQSEDSAKTEBEVDKQWVE--BAKVAS 1151
Db 1639 FEHKSFTKLCKEHDGFTGEDSSAHLERKSDNPLDIAVTRADLERNIERIEDIAP 1698
Qy 1152 ALEKWKTAIREAQTFSRMHVLGMLDACIKWMSAENARCKVCPKKGDDKILCDECNK 1211
Db 1699 GLRVNRRALSEARSAAQVALCICQOLQKSIWEKSIWKVYVCICRKGNEELLLLCDGCDK 1758
Qy 1212 AFHLCLRLALYEPDGEWQCACQAPATA-----RNSRGRNTEESASE-----DS 1258
Db 1759 GCHTYCHRPKITIIPDGFWFCFACIAKASQTLKIKKLHVKGKTNESKKGKVTLTGDT 1818

Qy 1259 EDESDDEEBEEEEEDYEVAGLRLRPKRTIRGKHSVIPPAARSGRRPGKKPHSTRR 1318
Db 1819 EDEDS-----ASTSSSLKRGKDKQK 1841
Qy 1319 SQKAPPVDDAEVDELVLQTKSSRRQSLQKCEILHKIVKYRFSWFRBVPVTRDEAE 1378
Db 1842 MEENTS--INLSQESFTSVKKPKRDDSKDLALCSMLITEMETHEDAWPELLPVNLKLPV 1899
Qy 1379 DYVDVITHPMDFQTVQNKCSGYSRVSQEFELTDMKQOVETNAEVYN 1423
Db 1900 GYKWKIKRPMDFSTIREKLSGQYPNLETFALDVLVDFDNCETFN 1944

RESULT 14

US-09-839-479-21
; Sequence 21, Application US/09839479
; Patent No. 6727222
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042002
; CURRENT APPLICATION NUMBER: US/09/839,479
; PRIOR FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 21
; LENGTH: 1972
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-839-479-21

Query Match 6.7%; Score 533; DB 4; Length 1972;

Best Local Similarity 18.0%; Pred. No. 2,1e-29;

Matches 353; Conservative 256; Mismatches 598; Indels 758; Gaps 68;

Qy 8 KPFLVNVLPGEPEF-FTIPH-----TQAFRTREYVEARLERYSERIMTCKST 55
Db 189 KPLSVNQAKETYMKLIVPSDPVLKAGNKNTSESSLTSELRSKEQY-----KQA 241

Qy 56 GSSQLTHKEAWEEEOEVAELK-----BEFPAYEKLVLVEMVHN-----TA 97
Db 242 PPSQLKKQESSKSLKVVIAALSNPKATSSSPA-HPKQTLNHNHPFLTNALLGNHPNG 300

Qy 98 SLEKLVDTAWLEIMTKYAVGECDFEV----- 124
Db 301 VIQSVIQEAPLALTATTKMQSKINENIAAASSTPSPVNLSTSGRRTPGNQTPVMPAS 360

Qy 125 -----GKEKML--KVKLVKI---HPL-----EKVDEE 146
Db 361 PILHQGKEKAVNNVNPVKQHHSHPAKSLVQFRGTSDIPSSKDSDEDEEEDDE 420

Qy 147 ATEKSDGACDSPSSDKENSSQIAQHQKQKTVVKEDEGRRESINDRARRSPKLPSTSLK 206
Db 421 EBEDEDEDEDDSDSQSDESDSNTSESDTEGSEEDDDDDQDDESD-----T 467

Qy 207 KGERKWAAPPKFLPHKYDVKLQNEDKIINV--PADSLIRTERPPNKEIVRYFIRNALRA 264
Db 468 EGE-----KTSMKL---NKTSSVKSPSMLTGHSTPRNLHAK----- 503

Qy 265 GTGENAPWVVEDELVKYKSLPSKFSDFLLDPKYKMTLNPSTKTKNTGSPDRKPSKSK-T 323
Db 504 -----APGSAAPALCSESQSPA-----FLGTSSSTL---TSSPHSGTSKRRRV 544

Qy 324 DNSSLSSPLNPKLWCHVHLKKSLSGSLPKVKNKSNKSP-----EBHLEMMIOMSPN---- 376


```
Db 779 LBERQKQMIIEEMKXPTEDMCLTDHQPDPFSRV---PGLTLPSCAFSDCLTIVEFLHS 835
Qy 620 YSGLLPDAQYPTITAVSLMEALSADKGFYLNRVLVILLQTLLOLQDEIAEDYGE-- 677
Db 836 PGKVLGDPADKDVPSLGVLCQGLCCGDSL---GEVQDLLVRLKAAALDPGPPSYCQSK 892
Qy 678 --LGMKLSIPIPLTHSVSELVRLCLRRSDVQSESGSDTDNDKDSAAFEDNEVDQEFLEK 735
Db 893 KILGEKVSIPLTRDNVSEILRCFLMAYG-----VEPALCDR 929
Qy 736 LETSBEFFELTSEKLIQILTALCHRI---LMTYSVQDHMETRQMSAELW--KERLAVLK 789
Db 930 LRTQPPQAQPPQKAAVLAFFVHELNGSTLIINEIDKTLESMSYKKNKWIVEGRULRLK 989
Qy 790 BENDKKRAEKOKRKEKEMAKNKGKVENGLGKTDRKKRIVKFEPPQVDTEAEDMISAVKSR 849
Db 990 ----TVLAKRTGRSEVEM-----GRPECLGRRSSRIEETSGMEESEESIAAVPCR 1040
Qy 850 RLLAIQAKKERIEQEREMKV-KLERQABEERIRKHKAAAEKAFQEGIAKAKLVMTPTIG 908
Db 1041 R-----GRDGEVDATASSIPELERQIEKLSKR-----QLFPRKLLHSSQMLRAVSLG 1089
Qy 909 TDRNHNRYW-----LFSDEV-----PGLFIEK----- 930
Db 1090 QDRYRRRYWVLYLAGIFVEGTEGNLVPEEVIKETDLSKVAHAASLNPAFLFSMKMELAG 1149
Qy 931 ----- 930
Db 1150 SNTTASSPARASRLPKTKPFQMPREFKSPVRGQDSEQPAQLOPEAQHVPAPQPOL 1209
Qy 931 -----GWV-----HDSI----- 937
Db 1210 QLQLOSHKGLEQSGPLSIGOSQHDLSQSAFLMSLQTHSSLSLSSSVLTDPDSSPGKLD 1269
Qy 938 -----DYRENHCK-----DHTVSGDEDCPRSKKANLG 966
Db 1270 PAPSQPPEPEDEABESSPDLOAFWENISAQMPCNAAPTPLAVSSEQ---PTPSPQOLA 1326
Qy 967 KNASMN----- 972
Db 1327 SSKPMNRPSAANPCSPVQFSSTPLAGLAPKRRAGDPGEMQSPQTGLQPKRRGRPPSKFF 1386
Qy 973 ---TQHGTAEVAVETTPKQGNLWFLCDSQKELDELLNCLHPGIRBSQLKERLEKYQ 1030
Db 1387 KQMEORVLTLQAQPVPEMCSGWMWIPDPPEM-LOAMLKALHPRGIREKALHKLHNK-HR 1444
Qy 1031 DLIHSIHLARKNLGLKSCDGNQELNLFRLSLIEVATEL---QKG-----G 1074
Db 1445 DFLQEVCL--RPS-----ADPIPEPQLPAPQEGIMSWSPKEXTYETD 1485
Qy 1075 LGYVBEETSFEARVIS-----LEKLKDFQECV----- 1101
Db 1486 LAVLQWVELEQVRVMSDLQIRGWTCPSPDSTREDLAYCEHLSDSQEDITWRGPGREGLA 1545
Qy 1102 -----IALQASVKKPLQGFMAPKQ---KRRKLOQESDA---KTEEVD 1138
Db 1546 POKRTTNPLDLAVMLRAALEQNVKRYLREPLWPTHEWVLEKALLSTPNGAPEGITTEIS 1605
Qy 1139 EEKQWVEEAKVASALEKWTAREACTFSRMHVLLGMLDACIKWMSAENACKVCYCPKG 1198
Db 1606 YE-----ITPRIRIMQTLQRCRSAHVCLCLGHLERSIAWEKSNVKNVTCLVCRKGD 1657
Qy 1199 EBDKILILDCNKAFLCLRLPALVEVPDGEWQCPACQAPATARRNSRGRNYTEESASEDS 1258
Db 1658 NDEFLLLCDCGRGCHYCHRPKMEAVPEGDMFCTVC-----LAQOV 1699
Qy 1259 EDESDSEEBEESBEEBDEYV---AGLRLPRKTIROKHSVIPPAAASGRPRPKKPHS 1315
Db 1700 EGFEQTKPGFPKRGOKRGKSGYSINTFSEGDGRRRRVLLKGRES---PAA-----GPRYSE 1750
Qy 1316 TRSQPKAPPVDDAEVDELVLQTKSSRRSQSLQELQCEILHKIVKYRFSWFPFVTRD 1375
Db 1751 ERLSPSKRRPL-----SMRNHSDLTFCETIILMEMESHDAAMPFLFEPVNP 1796
```

```
Qy 1376 EAEDYDVITHPMDFTQVQNKSCGYSRSVQBFLLTDMKQVFTNAEYVNCRGSHV 1429
Db 1797 LVSGYRRIIKNPMDFSTMRERLLRGYTSSEFAADALLVFDNCQTFNEDDSEV 1850

RESULT 16
US-09-839-479-70
; Sequence 70, Application US/09839479
; Patent No. 6727222
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042002
; CURRENT APPLICATION NUMBER: US/09/839,479
; CURRENT FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 70
; LENGTH: 1876
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-839-479-70
```

```
Query Match 6.5%; Score 518.5; DB 4; Length 1876;
Best Local Similarity 19.4%; Pred. No. 2.2e-28;
Matches 267; Conservative 177; Mismatches 429; Indels 501; Gaps 44;
```

```
Qy 415 QKSTGNKSPKGLKPTKTKMOMTLLDMAKGTQ---KMTAPRNSGGTPTSSKPHKHL 471
Db 619 QAITGKRGRPR---NTERAKTKEVPKVRGRPRPKVITEL-----LNKTDNRPLKKL 669
Qy 472 PPAALHLATYYKENKDRKRSALSCVISTKARTLSSSEDRARLPPELRSLVQKRYELLEH 531
Db 670 EA-----QETNEEDKAKIAKSKKMKQKQVORGECLTTIQOQARNKFQETKSLKH 720
Qy 532 KKRWSMSEBQEKYLLKKREELKKLKEKAKEREKE-----M 570
Db 721 KZ--AKKSKAKERKGTQKQELKEKVRKEKEKVEEYTKAKPACKADKTLATQRR 778
Qy 571 LERLEKQ-----KRYEDQELTG-KNLPAFLVDTPEG--LPNTLFGDVAMVVEFLSC 619
Db 779 LEERQKQMIIEEMKXPTEDMCLTDHQPDPFSRV---PGLTLPSCAFSDCLTIVEFLHS 835
Qy 620 YSGLLPDAQYPTITAVSLMEALSADKGFYLNRVLVILLQTLLOLQDEIAEDYGE-- 677
Db 836 PGKVLGDPADKDVPSLGVLCQGLCCGDSL---GEVQDLLVRLKAAALDPGPPSYCQSK 892
Qy 678 --LGMKLSIPIPLTHSVSELVRLCLRRSDVQSESGSDTDNDKDSAAFEDNEVDQEFLEK 735
Db 893 KILGEKVSIPLTRDNVSEILRCFLMAYG-----VEPALCDR 929
Qy 736 LETSBEFFELTSEKLIQILTALCHRI---LMTYSVQDHMETRQMSAELW--KERLAVLK 789
Db 930 LRTQPPQAQPPQKAAVLAFFVHELNGSTLIINEIDKTLESMSYKKNKWIVEGRULRLK 989
Qy 790 BENDKKRAEKOKRKEKEMAKNKGKVENGLGKTDRKKRIVKFEPPQVDTEAEDMISAVKSR 849
Db 990 ----TVLAKRTGRSEVEM-----GRPECLGRRSSRIEETSGMEESEESIAAVPCR 1040
Qy 850 RLLAIQAKKERIEQEREMKV-KLERQABEERIRKHKAAAEKAFQEGIAKAKLVMTPTIG 908
Db 1041 R-----GRDGEVDATASSIPELERQIEKLSKR-----QLFPRKLLHSSQMLRAVSLG 1089
Qy 909 TDRNHNRYW-----LFSDEV-----PGLFIEK----- 930
```

```

; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 1878
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(1878)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-09-418-710-13

Query Match      6 4%; Score 512.5; DB 4; Length 1878;
Best Local Similarity 19.3%; Pred. No. 6.1e-28;
Matches 263; Conservative 179; Mismatches 443; Indels 479; Gaps 42;

Db      QY      1090 QQRYYRWVLPYLAGIFVEGTGELNVEEVIKKETDSLKVAHAASLNPAFLFSKQELAG 1149
          931      -----
Db      QY      1150 SNTTASSPARARSRLTKPGMPQPREFKSPVRGQDSQPOQAQLHVPAPQPQL 1209
          931      -----
Db      QY      1210 QLQLOSHKGFLEQSGPLSGOSQHDLSQSAPFLSLSQTSQSHSSLLSSSVLTPDSSPGKLD 1269
          938      -----
Db      QY      1270 PAPSPPEPEDEAESPDLOAFWFIISAQMPCNAAPTPLAVSEDO---PTSPQOLA 1326
          967 KNASMN-----
Db      QY      1327 SSKPNRPSAANPCSPVQFSSTPLAGLAPKRRAGDPGEMPSQPTGLGQPKRRGRPPSKFF 1386
          973 ---TQGTATEVAVETTTKQGNLWFLCDSQKELDELNCLHPQIRSSQLKERLEKRYQ 1030
          1387 KQMEQRLVTQTAQPVPPGEMSGWWIIPDEM-LDAMLKALHPRIREKALHKLHNK-HR 1444
          1031 DIIHSIHLARKENGLKSCDGNQELLNPLRSOLIEVATRL---QKG-----G 1074
          1445 DFLQEVCL--RPS-----ADPIPEPQLPAFGIWSGPKETTYETD 1485
          1075 LGYVEETSEFEARVIS-----LEKLKDFGECV----- 1101
          1486 LAVLQWVELEQQRVIMSDLIQIRGWTCPSPDSTREDLAYCEHLSDSQEDITWRGPGREGLA 1545
          1102 -----IALQASVIKKFLQGFMAPKQ-----KRRKLOSDSA---KTEEVD 1138
          1546 POKTTNPLDLAVMRLAALAEQNVKRYLRLEPLPHEWVLEKALLSTENGAGEGTTTBS 1605
          1139 EKKQWSEKAVASALEKKWTAIREAQTSRHHVLGLMDACIKWDMSAENARCKVCPKKG 1198
          1606 YE-----ITPRIRIWRQTLQRCRSAHVCLGLHLERSTAMEKSVNKTCLVCRKGD 1657
          1199 EDDKLILDECNKAFHFLCLRPALYVDPGEWQCPCAPATARRNSRGRNTEESASBDS 1258
          1658 NDEFLLCDGDRGCHYCHRPKMEAVPEGDFCTVC-----LAQOV 1699
          1259 EDDSEDESEEEEEEEEDVEV-----AGLRRLRPRTIRKGSVIPPAAARGRRRPPKPKHS 1315
          1700 EGEFTQKPGPKRGOKRSGSYSLNFSQDGRRRRLVLLKGRS---PAA-----GPRYSE 1750
          1316 TRSQPKAPPVDDAEVDLVLQTKESSRRQSLQKCEIHLKIVKRYFSMPFERPVTRD 1375
          1751 ERLSPSKRRPL-----SMRNHSDLTFCIILMEMESHDAAMPFLPEPVNPR 1796
          1376 EAEDYVDVITHPMDFTQVQNKSCGSYSRVSOFELTDMKQVFTNARVYVNCGRSHV 1429
          1797 LVSGYRRIIKNPMDFSTMRERLLRGYTSSEFAADALLVFDNCQTFNEDDSEV 1850
```

RESULT 17

```

US-09-418-710-13
; Sequence 13, Application US/09418710
; Patent No. 6596482
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042001
; CURRENT APPLICATION NUMBER: US/09/418,710
; CURRENT FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 73
```

Db 1391 MEQRYLTQTAQVPPVPMCSGWWIPDPEDMLDAMLKALHPRGIREKALHKLHNLK-HRDFL 1449
Qy 1034 HSIHL--ARKPNLGLKSCDGNQF--LLNFLRSLDIIEVATRLQKGLGVBEETSFEARVIS 1090
Db 1450 QEVCLRPSADPIFEPRQLPAFGGIMSWSPKE-----KTYETDLAVLQWVELEQRVIM 1503
Qy 1091 -----LEKLKDFGECV----- 1101
Db 1504 SDLIQIRGWTCPSPDSTREDLAYCEHLSDSQEDITWRGPGREGIAPQRTTNPLDLAVMRL 1563
Qy 1102 IALQASVIKKFLGFWAPQKRKLQSDSASKEEVEDEEKQWVE-----EAK 1148
Db 1564 AALEQNVKRYLREPLWP-----THEVLEKALLSTPNGAPEGTTTTSYE 1609
Qy 1149 VASALEKWKTAIREAQTFSRMHVLLGMLDACIKWMSAENARCKVCPKKGDDKJLILCDE 1208
Db 1610 ITPRIRIWRQTLQRCSAAHVCLIGHLERSIAWKSVNKVTCLVCRKGDNDFLLCDG 1669
Qy 1209 CNKAFHLCLRPALYEVDPGEWQCPACQAPATARRNSGRNYTEESASESDEDEEBEE 1268
Db 1670 CDRGCHYCHRPKMEAVPEGDFCTVC-----LAQQVEGEFTQKPGF 1711
Qy 1269 EEEEEEEEDYEV--AGLRPRKTIIRGKHSVIPPAARSGRPPGKPHSTRSQKAPP 1325
Db 1712 PRKGQKRGSGYSLNFSEGGRRRRVLLKGRES---PAA--GPR-----YSEERLSPSK-- 1759
Qy 1326 VDDAEVDELVLQTKRSRRQSLQKCEELHKIVKYSWPPREPVRDEADYDVIT 1385
Db 1760 -----RRRLSMRNHSHDLTCEIILMEMESHDAWPXPVNPRLVSGYRRIIK 1808
Qy 1386 HPMDFTQVONKSCGSYRSVQBEPLTDMKQVFTNAEYVNCRGSHV 1429
Db 1809 NPMDFTMRERLLRGYTSSEFSAADALLVDFNCQTFNEDDSEV 1852

RESULT 18

US-09-839-479-13
; Sequence 13, Application US/09839479
; Patent No. 6727222
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042002
; CURRENT APPLICATION NUMBER: US/09/839,479
; CURRENT FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 1878
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(1878)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-09-839-479-13

Query Match 6.4%; Score 512.5; DB 4; Length 1878;

Best Local Similarity 19.3%; Pred. No. 6.1e-28;

Matches 263; Conservative 179; Mismatches 443; Indels 479; Gaps 42;

Qy 415 OKSTGNSKSPKGLTKPTKMKQMTLLDMAKGTQ---KMTAPRNSGGTPTRTSSKPKHL 471

Db 619 QAITGRGRPR---NTEKAKTEKVEPKVKGRRGPPVKITEL-----LNKTDNRPLKKL 669

Qy 472 PPAALHLIAYYKENDKREDKRSALSVCISKARTALSSSEDRARLPBELRSIVQKRYELLEH 531
Db 670 EA-----QETLNEEDKAKIAKSKKMRQKVQGECLTTIQGOARNKRKQETSKLKH 720
Qy 532 KRWASMSSEQRKEYLKKREBELKKLKEKAKER---REKE----- 569
Db 721 KE--AKKKSXAEBKKGTKQEKLEKVKREKKEKVQKEKEEVTAKPACAKADKTILATOR 778
Qy 570 MLERLEKQ-----KRYEDQELTG--KNLPAPRLVDTPG--LPNTLFGDVAMVVEFLS 618
Db 779 RLEERQKQOMILEEMKKPTEDMCLTDHQPDPFSRV---PGLTLPSSGASDCLITIVFFLH 835
Qy 619 CYSGLLLPDAQYPIITAVS--LMEALSADKGGFYLYNRVLVILQTLQTLQDIAEDYGE 677
Db 836 SFGKVLGFPAPKDVPSLGVLCQGLLCOGSLGSEVQDLLVRLKAAALHDGFPSPYCQSLKI 895
Qy 678 LGMKLSIEPLTILHSVSELVRLCLRSDDVOEESGSDTDNKSAAAFEDNEVQDEFLEKLE 737
Db 896 LGKRVSEIPLTRDNVSEILRCFLMAYG-----VXPALCDRLR 932
Qy 738 TSSEFPELTSEKQLQILTALCHRI---LMTYSVQDHMETRQOMSABLW--KERLAVLKEE 791
Db 933 TOPFOAQPPQQAVALVAFPVHELNGSTLIINEIDKTLESMSYRKNKWIVEGLRLRK-- 990
Qy 792 NDKKRAEKOKRMEKAKNKENGVENGLKTKRKRIVKPEPQVDTEAEDMISAVKSRRL 851
Db 991 --TVLAKRTGRSEVEM---GRPECLGRRSSRIMBETSGWEEEEEESIAAIVPGR-- 1042
Qy 852 LAIQAQKEREIOEREMKV--KLBRQAEEERIRKHAAAEEKAFQEGIAKAKLWMRTPIGTD 910
Db 1043 ---GRRDGEVDATASSIPELERQIEKLSK---QLFFRKLKLSHSSQMLRAVSLGQD 1092
Qy 911 RNHNRYWLFSDVPGLFIB-----KGWVHDSID----- 938
Db 1093 YRIRRYWVL--PYLAGIFVETGTEGNLVPEEVIKKTDSLVAAHASLNPALFSWKMBLAGS 1151
Qy 939 ---YRFNHCKDHTYSGDEDCPRSK----- 961
Db 1152 NTTASSPARARSRLKTKPGFQPRHFKS--PVRGQDSEQFQAQLQPEALQHLVPAQPQL 1210
Qy 962 ---KANLGN-----ASMTQHGTTATEVAVETTPKQG-- 991
Db 1211 QLQLQSHKGFLQEGSPLSGOSQHDLSQSASFLSMLSQTSQSHSLSSSVLTDPDPSGKL 1270
Qy 992 ---QNLWFLCDSQ----- 1001
Db 1271 DPAPSQPPEEPDEAESSPDLOAFWFNISAQMPCNAAPTPLAVSEDOPTSPSPOOLASS 1330
Qy 1002 --- 1001
Db 1331 KPMNRPSAANPCSPVQFSSTPLAGLAPKRRAGDPGEMPOSPTGLQPKRRGRPPSKFFKQ 1390
Qy 1002 ---KELDELLNCLHPQIGIRESQKLERLEKRYQDII 1033
Db 1391 MEQRYLTQTAQVPPVPMCSGWWIPDPEDMLDAMLKALHPRGIREKALHKLHNLK-HRDFL 1449
Qy 1034 HSIHL--ARKPNLGLKSCDGNQF--LLNFLRSLDIIEVATRLQKGLGVBEETSFEARVIS 1090
Db 1450 QEVCLRPSADPIFEPRQLPAFGGIMSWSPKE-----KTYETDLAVLQWVELEQRVIM 1503
Qy 1091 -----LEKLKDFGECV----- 1101
Db 1504 SDLIQIRGWTCPSPDSTREDLAYCEHLSDSQEDITWRGPGREGIAPQRTTNPLDLAVMRL 1563
Qy 1102 IALQASVIKKFLGFWAPQKRKLQSDSASKEEVEDEEKQWVE-----EAK 1148
Db 1564 AALEQNVKRYLREPLWP-----THEVLEKALLSTPNGAPEGTTTTSYE 1609
Qy 1149 VASALEKWKTAIREAQTFSRMHVLLGMLDACIKWMSAENARCKVCPKKGDDKJLILCDE 1208
Db 1610 ITPRIRIWRQTLQRCSAAHVCLIGHLERSIAWKSVNKVTCLVCRKGDNDFLLCDG 1669
Qy 1209 CNKAFHLCLRPALYEVDPGEWQCPACQAPATARRNSGRNYTEESASESDEDEEBEE 1268

Db	1670	CDRGCHTYCHRPKMEAVPEGDWFCTVC-----LAQQVEGEFTQKPGF	1711
Qy	1269	EEEEEEEEEDYEV---AGLRPRKTRTGRKHSVIPPAARSGRRPGKKPHSTRSQPKAPP	1325
Db	1712	PKRQKRGKSGYSLNFSFGDGRRRVLLKGRES---PAA---GPR-----YSEERLSPSK--	1759
Qy	1326	VDDAEVDVLQTKSSRSRSLQKCEEILHKIVKVPFSPWPPEPVTTRDEADYDVIT	1385
Db	1760	-----RRRLSRNRNHSDLTFCEIILMEMESHDAAPFPXPNPLVSGYRIIK	1808
Qy	1386	HPMDFQIVQNKSCGSGYSRVQVEFLTDMMKVFTNAEVNCRGSHV	1429
Db	1809	NPMDFTSMRRLRLLRGYTSSEEFADALLVFDNCQTFENDDSEV	1852

```

RESULT 19
US-09-418-710-72
; Sequence 72, Application US/09418710
; Patent No. 6596482
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042001
; CURRENT APPLICATION NUMBER: US/09/418,710
; CURRENT FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 72
; LENGTH: 1969
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-418-710-72

```

Query Match	6.4%;	Score 509.5;	DB 4;	Length 1969;
Best Local Similarity	17.9%;	Pred. No. 1.1e-27;		
Matches 351;	Conservative 247;	Mismatches 604;	Indels 763;	Gaps 67;
QY	8	KPFPVLNPLPCEBPF-FTIHP-----TQAFRTREYEARLYSERIWTCKST	55	
Db	189	KPLSLVNQAKETTYKLVIPSPDVLKAGNKNTSESSLTSELRSKEQY	241	
QY	56	GSSQLTHKEAWEEQEVALLK-----EEPPAWYEKLVLEMVHNN-----TA	97	
Db	242	FPSQLKKQESSKSLKVKVIAALSNPKATSSSPA-HPKQTLNHNPPFLTNALGNHPNG	300	
QY	98	SLEKLVDATAMLEIMTKYAVGEEDCFEV-----	124	
Db	301	VIQSVIQEAPLALTKTKYQMSKINENIAAASSTPFSSPVNLSTSGRTPGNQTWPMSAS	360	
QY	125	-----GKEKML-----KVIVKI-----HPL-----EKVDE	145	
Db	361	PILHSQCKEKAVSNVNPVPKTOHHSHPAKSLVEQFRGTDSDIPSSCKDSNDEEEDDE	420	
QY	146	EATEKKSGDAGDSPSSDKENSQIADHQKQETVVKEDEGRRESINDRRSPRKLITSL	205	
Db	421	EEDDEEDDESDDQSQSDSNSE--SSTEGSEEDDDKKQOQSDSD--	465	
QY	206	KKGERKWAPPKFLPHKYDVKLQNEDKIISNPADSLITERPPPNKEIVRYFIHNALRAG	265	
Db	466	TEGE-----KTSMKLNTKTS--SKSPMSLTHGSHSTPNLHIAK-----	501	
QY	266	TGENAPWVVEDELVKYKSLPSKFSDFLLDPKYMTLNPSTKRNKNTGSPDRKPSSKSK-TD	324	
Db	502	-----APGSAPALCSQSQSPA-----FLGTSSSTL--TSFPHSGTSKRRRVTD	543	
QY	325	NSLSLSSPLNPKLWCHVLKKSLSGSPVKVKNKSKSP-----EEHLEEMKMSFN----	376	

Db	1537	LRGIREKALQKQIQK-HLDYITOACLKNK-DVAIIELNENE--NQVTRDIVENVSVEEQ	1592
Qy	1070	LQKGGLGYVBETSEFEARVTSLEKIKDFGECVIALQASVIKKFLOGFMAKQ-----	1121
Db	1593	AMWMDLSVLQOVEDLERVAS-----ASLQ-----VKGMCEPASEREDLV	1634
Qy	1122	-----KRRKQSDSASAKTBEVDEBKWVE--BAKVA	1150
Db	1635	YFEHKSFTKLCXHDGFTGEDSSAHALERSKDNPLDIATVRLADLERNIERRIEDIA	1694
Qy	1151	SALEKKWKAJREAQTFGRMHVLLGMLDACIKWMSAENARCKVCPKKGEDDKLILCDECN	1210
Db	1695	PGLRVRRALSEARSAAQVALCIIQLOKSIWEKSIKVVYQICRKGDNELLLLCDGCD	1754
Qy	1211	KAFHLFCLRPALYEVVPGEWOCACQPATA-----RRNSGRNYYTESASE-----D	1257
Db	1755	KGCHTYCHRPKIIITIPDGNFCFACIIKASQOTLKIKKLHVKGKKTWESKKGKVTLTGD	1814
Qy	1258	SEDDSEDEEEEEEEEDYEVAGURLRPRKTIROKHSVIIPPAARSGRRPKKPHSTR	1317
Db	1815	TEDEDS-----ASTSSSLKRGNKDLOKQR	1837
Qy	1318	RSQPKAPVDDAEVDELVLQTKSSRRQSLELOKCEBIIHKIVKYPSPWFRFPVTRDEA	1377
Db	1838	KMEENTS--INLSKQESFTSVKPKRDSKDLCALCSMILTEMETHEDAWPFLLPVNLKLV	1895
Qy	1378	EDYYDVITWHPDQTQVONKSCSGSVRSVQEFELTDMKQVFTNAEYV	1422
Db	1896	PGYKKVTKKMDFTSIREKLSSGQYPNLETFALDVLRLVFNFCSTF	1940

RESULT 20

```

US-09-839-479-71
; Sequence 71, Application US/09839479
; Patent No. 6727222
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042002
; CURRENT APPLICATION NUMBER: US/09/839,479
; CURRENT FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 71
; LENGTH: 1969
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-839-479-71

```

```

Query Match      6.4%; Score 509.5; DB 4; Length 1969;
Best Local Similarity 17.9%; Pred. No. 1.1e-27;
Matches 351; Conservative 247; Mismatches 604; Indels 763; Gaps 67;

QY      8 KPFFLVNLPGEPP-FTIPH-----TQEAFTTREETEYEARLERYSERIWTCKST 55
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      189 KPLSLVNAQKETTMYLIVSPDVLKAGNKNTSESSLTSELRSKREQ-----KQA 241

QY      56 GSSQLTHKEAWEEBQEAELK-----EEFPATYEKLVLEMVHN-----TA 97
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      242 FPSQLKQESSKSLKKVIAALSNPKNTATSSSPA-HPKQTLNHNHPFLTNALGNHQPNG 300

QY      98 SLEKLVDTAMLEIMTKYAVGECEDFV-----V-----V-----V----- 124
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      301 VIQSVQEAFLALTKTKTQMSKINENTAAASSTPFSSPVNLTSGRRTPNGQTVPMPSPA 360

```

[illegible]

Db	1299	LSYONSGKSLGVSQVSTATQSNVEKADSNLFTNGSSGPGKFYPLPNDQLLKTLTBKNR	1358
Qy	992	QNLWFL-----CD-----	999
Db	1359	Q--WFSLLPRTPCDDTSLTHADKSTASLVTQPQSPKSPPTPAPLGSSAQNPVGLNPF	1416
Qy	1000	-----	999
Db	1417	ALSPLOKGVGVMGLQFCGWPTGVVTSNIPFTLSVPSLGLSGLSEBNGNSFLTNSVAS	1476
Qy	1000	-----SQKELDELLNCLH	1012
Db	1477	SKSESPVONEKATSAQAAAVEVAKPVDPSPKPIPEMOFGWRIIDPEDLKALLKVLH	1536
Qy	1013	POGIRESQLKERLEKRYODIHSHTLARKNGLKSCDGNQELLNFLRSOLIE---VATR	1069
Db	1537	LRGIREXALQKQTK-HLDVITQACLKNK-DVAIIELNENE--NQVTRDIVENWSVEEQ	1592
Qy	1070	LQKGLGYVEETSEFEARVLSLEKLDGFCBVIALQASVIKKFLQGFMAPKQ-----	1121
Db	1593	AMENDLSVLQOVEDLERRVAS-----ASLQ-----VKGMWCEPASEREDLV	1634
Qy	1122	-----KBRKLOSEDSAKTVEVDEEKKWVE--EAKVA	1150
Db	1635	YFEHKSTKLCKEHGDEFTGEDESSAHALERKSDNPLDIAVTRLADLERNIERREDIA	1694
Qy	1151	SALEKWKTAIREAQTFGRMHVLGMLDACIKWDMSAENARCKVCPKKGEDDKLILCDECN	1210
Db	1695	PGLVWRRLRAISEARSAOVALCTQLOQKSTANEKSIINKYVCQICRKGDNELLLLCDGCD	1754
Qy	1211	KAPHLFCLRPALVEVPDGEWQCPACOPATA-----RRNSRGRNYTESASE-----D	1257
Db	1755	KGCHTYCHRPKITTIPGDWFCPACIAKASQOTLKKLHVKGKKTNESKKGKVTLTGD	1814
Qy	1258	SEDDSEDEEEEEEEDEEYEVAGURLRPRKTIIRKSHSVIPPAABSGRRPKKKPHSTR	1317
Db	1815	TEDEDS-----ASTSSSLKRGKQLOQR	1837
Qy	1318	RSOPKAPPVDDAEVDELVLQTKRSSRRQSLELOKCEBILHKIKVYRFSWFFRPVTRDEA	1377
Db	1838	KMENTS--INLSKQESFTSVKPKPRDSDKDALCSMLITEMETHEDANFWFLIPVLKLV	1895
Qy	1378	EDYVDVITHPMDFOVQNKSCGYSYRQVBFLTDMKQVFTNAEYV	1422
Db	1896	PGYKKVTKKPMDFSTIREKLSGGOYPLNETFALDVLRFVFNCGTF	1940

```

RESULT 21
US-09-513-999C-7181
; Sequence 7181, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert, A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.

```

Query Match 4.4%; Score 352; DB 4; Length 60;
Best Local Similarity 100.0%; Pred. No. 1.9e-18;

```

Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 902 MRRTPIGTDNRNHRNYLFSDEVPGLFIEKGWVHDSIDYRFNHHCKDHTVSGDSDYCPRSK 961
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1 MRRTPIGTDNRNHRNYLFSDEVPGLFIEKGWVHDSIDYRFNHHCKDHTVSGDSDYCPRSK 60

RESULT 22
US-09-270-767-45282
; Sequence 45282, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 45282
; LENGTH: 800
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
; FEATURE:
; OTHER INFORMATION: Xaa means any amino acid
US-09-270-767-45282

```

RESULT 23
US-09-418-710-68
; Sequence 68, Application US/09418710
; Patent No. 6596482
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042001
; CURRENT APPLICATION NUMBER: US/09/418,710

; CURRENT FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 68
; LENGTH: 65
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-418-710-68

Query Match 3.9%; Score 311; DB 4; Length 65;
Best Local Similarity 98.5%; Pred. No. 2e-15; Indels 0; Gaps 0;
Matches 64; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 630 YPITAVSLMEALSADKGGFLYNRVLLVILLOTLLOLQDEIAEDYGLGMLSKIPLTL 689
Db 1 YPITAVSLMEALSADKGGFLYNRVLLVILLOTLLOLQDEIAEDYGLGMLSKIPLTL 60

QY 690 HVSVE 694
Db 61 HVSVE 65

RESULT 24
US-09-839-479-67
; Sequence 67, Application US/09839479
; Patent No. 672722
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042002
; CURRENT APPLICATION NUMBER: US/09/839,479
; PRIOR FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 67
; LENGTH: 65
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-839-479-67

Query Match 3.9%; Score 311; DB 4; Length 65;
Best Local Similarity 98.5%; Pred. No. 2e-15; Indels 0; Gaps 0;
Matches 64; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 630 YPITAVSLMEALSADKGGFLYNRVLLVILLOTLLOLQDEIAEDYGLGMLSKIPLTL 689
Db 1 YPITAVSLMEALSADKGGFLYNRVLLVILLOTLLOLQDEIAEDYGLGMLSKIPLTL 60

QY 690 HVSVE 694
Db 61 HVSVE 65

RESULT 25
US-09-949-016-11433
; Sequence 11433, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.

; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11433
; LENGTH: 2733
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-11433

Query Match 3.8%; Score 302.5; DB 4; Length 2733;
Best Local Similarity 18.5%; Pred. No. 1.7e-12; Indels 661; Gaps 80;
Matches 358; Conservative 279; Mismatches 633; Indels 661; Gaps 80;

QY 30 EAFRTREYEARELERYSERIWTCKSTG-----SSQLTHKEANWEEBOEVALLEKBEPPA-- 82
Db 505 ETERGEVEEDKENKEYSEKCVTSKQOEIYILKQITSEKEV--ELQHIRKLEEKLAABE 562

QY 83 WYEKLVLEM-----VHHTASLEKLV-----DTAWLEIMTKYAV 116
Db 563 QFQALVKQKQNTLQDKTQIDLLQAEISENQAIIOKLITSNTDSDGDSVAL-VKTVVI 621

QY 117 GEEDCFEVG-----KEKMLKVIKIHPLKLVDEEATEKKS--DGACDSPSDKEN 165
Db 622 SPPC---TGSSEHWKPELEEKILALEKEQKQKQLEALTSRKAILKKAQEKERHREE 678

QY 166 SSQIAQHQKQKTVVKEDEGRRESINDARRSPRLPTSLKKGKRWKAPKPLPHKYDKV 225
Db 679 LKQKDDYNRLQEQDEQSKENENICDQLRQIQVRESI-----DGK 721

QY 226 LQNEQ--KLIISNVA-----DSLIRTERPNKEIVYFIRHNALRAG 265
Db 722 LPSTQQSCSSTPGLBEPLFKATQHHTQPVLESNLCDFWPSHSEDA-----SALQGG 775

QY 266 TG-ENAPWVVEDELVKYSLSPKSFDFLDLPYKMTLNPSTPKRKNKTSRDRKPSKSKT 324
Db 776 TSVAQIKAKLKEIEAEKVELELVSS-----TSELTKSEEVFQLOQINKQGLE 826

QY 325 NSSLSPLNPKLWCHVH---LKSLSGSPLKVNKSNKSPPEHLEEMMMKSPNK----- 377
Db 827 IESLKTVSHE---AEVHAESLQQLLESSQIQIAGLEHLRELQKLDQKLSKKEEDVS 883

QY 378 -----LHNTNPHI PKKGPAPKPKGKSHDK----- 400
Db 884 YLSGQSEKBAALTKIQTEIIEQEDLIKALHTQLEMQAK-----EHDEKILQLOVEL 935

QY 401 -PLKAKGRSGILNQKSTGSK-----SPKKGLTKPTKMKQMTLLDMAGT-QKMTR 452
Db 936 CEMKQKPEIEGESRAKQIQKQLAALISKEALKENKSLQELSL---AGTTERLT 992

QY 453 APRNSGGTPTRTSSPKPHKLPALHLYAYKENKDR-----EDKRSALSCVISKT 502
Db 993 SLADVESQVSAQNK-----KDTVGLRLLQBERDKLITEMDRSLLENQSLSSCESLKL 1048

QY 503 ARLLSSEDRARLPEELRSL-----VQKRYELL-----EHKKRWA----- 536
Db 1049 ALEGLTEDKEKLKVEIESLKSIAESTEWQEKHKELQKEYEILQSYENVNSNEARIQH 1108

QY 537 ---SMSEORKEYLKKREELKKLKEKAKEREKEMLERLEKOKRY----- 580
Db 1109 VVEAVRQEKQELYGLRSTEAANKKTEQLQEAQEQEMEMKEMKPKFAKSKQOKILELEE 1168

QY 581 EDQELTGKNLPAPRLVDTPEGLPNTLFGD-----VAMVVEFLS-CYSGLL----- 624

Db 1462 CEMKQPEEIGERSRAQIQKQALISRKALKENKSLQBELSL---ARCTIERLTK 1518
Qy 453 APRNSGGTPTRTSKPHKHLPPAALHLIAYYKENKOR-----EDKRSALSCVISKT 502
Db 1519 SLADVESQVSAQNE---KDTVLGRLLALLQBERDKLITEMDRSLLENQSLSSCSLSKL 1574
Qy 503 ARLLSSDRARLPPEELRSL-----VQKRYELL---EHKRWAA-----536
Db 1575 ALEGLTEDREKLVKEITSLSSKSIABSTEWQEKHKELOKEYEYILLOSIVNSNEARIQH 1634
Qy 537 ---SMSEORKEYLKKREELKKLKEKAKERREKEMLEKQKRY-----580
Db 1635 VVEAVQEQELYGLKRLSTANKKETEKQLOEABQEWEMKEKQRFKAKSKOOKILELEE 1694
Qy 581 EQDELTKGNLPAPRLVDTPEGLPNTLFGD-----VAMVVEFLS-CYSGLL-----624
Db 1695 ENDRLRAEVHPA---GDTAKECETLSSNASMKEELERVQMEYETLSKKFKQSLMSEKDS 1751
Qy 625 -----LPDAQYPI-----632
Db 1752 LSEEVQDLHQIEDNVSKQANLEATEKHNDQNTVTEEGTQSIPTGETEQDSLSMSTRPTC 1811
Qy 633 ---TAVSLMEALSADKGGFLYLNRLVLLIQ-----TLQLTL-----666
Db 1812 SESVPSAKSANPAVSKDFSSHDEINNYLQOIDOLKERIAGLBEEKOKNKEFQOTLENEKN 1871
Qy 667 -LQDEIARDYGEIGKMLSEIP---LTLHSVSELVRLCLRRSDVQSESGSDTDNNDKSA 721
Db 1872 TLLSQISTKDGELKMLQEEVTKWNLNQIQIELSRVTKLKTAEBEKD--DLERLMNQ 1929
Qy 722 AFE-----DNEVQDEFLKLETSFFEL-----TSEKQIILFALCHRILM 762
Db 1930 LAELNGSIGNYCODVTDQAIKNELLE---SEMKNLKKCVSELEBEKQOL-----1975
Qy 763 TYSVQDHMETROQMSAELKERLAVL---KEENDKKRAE-----KQKREMEAKNKEN 812
Db 1976 -----VKEKTVSEIRKYLEKIQGAKQPGNKSHAKELQELLKEKQEVKQLOKDC 2028
Qy 813 GKVENGLGTRDKRKIVKF---EPQVDTE--AEDMISAVKSR-----849
Db 2029 IRYQEKISALERTVKALEFVQIESQKDLITKENLAQAVEHRRKKAQAEALASFVLLDDTQ 2088
Qy 850 -----RLLAIOAKERIOEREMKV-----LER---QAEERIRKHAAAEKAPQE 893
Db 2089 SEAAVLADNLKXKELQSNKESVSKQKQEDLERRLEQAEKHLKEKNQMKEL--D 2146
Qy 894 GIAKAKLVMRPTPIG-----TDNRNRYWLFSDVPGFLPIEKGWTH 934
Db 2147 ALRREKVVHEET-IGEIQVTLNKKQKEVQOQLQENLDSTVTQLAAFTKSMSSLDQDRDVI 2205
Qy 935 DSI---DYRFNH--HCKDHTVSGDEDYCPRSK-----KANLGK---NASMN 972
Db 2206 DEAKKWERKESDAIQSEKEIRLKEDNCVLDQLRQMSIHMEELKINISRLHDKQIWE 2265
Qy 973 TQHGTADEVAVTTTPKQONLWFLC-----DSQKELDELINCLHPOGIRESQLK 1022
Db 2266 SQAQTEVOLQKQVCDTLQGENKELLSQLEBTRHLYHSSQNELAKL-----ESELK 2315
Qy 1023 ERLEKRYQDIHSHILARPKNLGLSKDCGNQEL-----INFLRSDI---LEVATRLQ 1071
Db 2316 S-LKQDLTDLNSLEKCKQKQNLGIIIRQGEADIQNSKFSYEOLFTDLQASRELTSLRH 2374
Qy 1072 KGLGLGVSETSFEARVLSLEKLDQFGEVIALQASVIKFLQGFMAPKQKRRKQSEDS 1131
Db 2375 -----EINKKEQKIILLSGKE-----EAIQVAIELRQOHDKYEIKELNLSQEBE 2422
Qy 1132 AKTEEVDEEKQWVEAKVASALEKWKTAIRE-----AQTFSRMVLGLMLDACIKWMSA 1186
Db 2423 ENIVLEENKKAVD--KTNLQMETLTKTKKENIQKQAK-----LDSFVKSMSL 2469
Qy 1187 ENARCKVCPKGEDDKLILCDECNKAFHLFCURPALYVEPDGEWQCPACQATARNRSG 1246
Db 2470 QNDR-----DRIV-----GDYQ-----2481

Qy 1247 RNYTEESASEDDDEDEEEEEEDEEYEVAGRLRPRKRTIRGKSHVPPAARSG 1306
Db 2482 -QLEERHLSIILEKQQLIQEAAAANNKKEE-----IRGLRSHMDDL---2522
Qy 1307 RRPGRKPHSTRSQPKAPVDDAEVDDELVLOTKRSSRRQSLE--LOKCEBILHLKIVKRYF 1364
Db 2523 -----NSNAKLDALIQYREDLNQVITIKDSQKQLLEVQLOONKELENYAKLEE 2574
Qy 1365 SWPFREPVTREAEADYVDVITHPMDFQTVQNKSCGYSYRVOEFLTDMKQVFTNAEVTNC 1424
Db 2575 KUKSEEEANEDLRFSNALQEBEKDLS-----KEISLKVSIQOLTRQVTALQE 2623
Qy 1425 RGS-----HVLSCMVKTEQCLVVLVHLKHL-----PGHPYVRRKRG 1459
Db 2624 EGTGLGYHAQLKVKKEEVHRLSALFSSSQKRIAELEELVVCQKEAAKKVGE--IEDKLG 2681
Qy 1460 KFPDRLADEG-----DSEPAVQO-SRD-----EDRSRAEAEQEWLQDTS 1500
Db 2682 KELKHLHDAGIMRNDETETAERVAELARDLVEMEOKLLMVTKENKGLTAQIQSGFRSMS 2741
Qy 1501 LYSAKINSKDH 1511
Db 2742 SLQ-----NSRDH 2749

RESULT 27

US-09-538-092-1131
; Sequence 1131, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
; APPLICANT: Giot, Loic
; APPLICANT: Mansfield, Traci A.
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538,092
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CurapatSeqformatter Version 0.9
; SEQ ID NO 1131
; LENGTH: 2375
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: Polypeptide Accession Number P46100
US-09-538-092-1131

Query Match 3.7%; Score 299; DB 4; Length 2375;

Best Local Similarity 18.6%; Pred. No. 2.5e-12;

Matches 312; Conservative 234; Mismatches 600; Indels 532; Gaps 67;

Qy 46 SRIWTCKSTGSSQLTHKEAWEEQVAVALLKEEFPANWEKLVL--EMVH-----HNTA 97
Db 188 SNKVEHTSRFSPKTSSNCNGEEKKJDDSCSGSVTYSYSAIVPKEMIKKAKLIETTA 247
Qy 98 S-----LEKLVDTAWLEIMTK-----YAVGECDFEV-GKEKML 130
Db 248 NNNSSVYVFLKQATDINSISSATKRLQKAFKSVLADIKKAHLAEEDLNSFRANDAVN 307
Qy 131 KVKIYKIHLEKVEATEKKSQDAGD-----SPSSDKENSSQIAQDHQKQKTVVKEDE 184
Db 308 KEKNTKEHKVIDAKETKARKEKPCALEKKDKISSEAKLSRKQVDSHM-HQNVPTREQ 366
Qy 185 GRESINDRARRSPKRLTSLKKGKRWAP-----PKFLPHKVDVQLONED 230
Db 367 RTNKGSTGGEHKSKDR-----KEEPQEPAPNTSDDLMDIVSVSPSPVEDIFENLETAM 419

Qy	229	EDKIISNV	PADSLIRTER	PNKEIVR	FI	RHNAL	RAC	TGENA	PVW	VEDEL	VKKYS	LP	SKF	288	
Db	960	EE	-----	EEAAR	KQLO	---	LEK	VTAE	AKI	KKLE	BD	DL	VMD	QNSKL	997
Qy	289	S	---	DFLLD	PV	KYMT	PLN	SP	TGR	KNTG	SP	DR	KPS	345	
Db	998	SKER	KL	BE	RVS	DU	T	T	W	LA	EE	E	K	1046	
Qy	346	LSG	---	SPUK	V	KN	S	KN	S	K	S	P	E	401	
Db	1047	EKS	Q	LE	K	KL	R	KG	D	AS	D	F	HE	1102	
Qy	402	LK	AGRS	-	KGI	---	LNQ	K	T	SG	N	S	K	442	
Db	1103	IAQ	NN	AL	K	K	I	R	E	L	G	H	S	1160	
Qy	443	MAK	T	Q	W	T	R	A	P	N	S	G	T	495	
Db	1161	-	STAT	Q	EL	RA	K	R	A	K	R	E	V	1218	
Qy	496	SC	V	I	K	T	A	R	L	S	S	D	R	555	
Db	1219	KAN	L	D	K	S	K	O	T	L	E	K	E	1265	
Qy	556	K	L	K	E	K	A	K	E	R	E	K	E	615	
Db	1266	SK	C	S	D	---	GER	A	R	A	E	L	S	1315	
Qy	616	FL	S	C	Y	G	L	L	P	D	A	Q	P	674	
Db	1316	Q	L	O	T	B	E	L	L	O	B	E	T	1356	
Qy	675	---	Y	G	E	L	M	K	L	S	E	I	P	724	
Db	1357	E	A	K	N	L	R	H	V	S	T	L	N	1416	
Qy	725	---	D	N	E	V	O	D	E	F	---	---	---	763	
Db	1417	K	L	E	K	T	K	N	L	Q	E	L	D	1466	
Qy	764	Y	S	V	O	D	H	E	T	Q	O	S	A	E	813
Db	1467	Y	A	---	---	DER	D	R	A	E	A	E	R	E	1521
Qy	814	K	V	E	N	L	G	K	T	D	R	K	B	I	861
Db	1522	D	V	G	K	N	V	H	E	K	S	K	R	A	1579
Qy	862	I	Q	E	R	---	---	EM	K	V	K	L	R	Q	908
Db	1580	L	O	A	R	D	E	Q	N	E	E	K	R	O	1639
Qy	909	T	D	R	N	H	N	Y	L	F	S	D	E	V	965
Db	1640	R	E	---	---	E	A	I	K	L	R	K	L	Q	1684
Qy	966	G	K	N	A	M	N	T	O	H	G	T	A	T	1012
Db	1685	L	E	A	D	M	L	Q	E	D	L	A	A	E	1735
Qy	1013	P	O	G	I	R	S	O	L	K	E	R	L	E	1072
Db	1736	---	A	R	I	A	Q	L	E	E	E	E	---	1763	
Qy	1073	G	G	L	Y	V	E	B	E	T	S	E	F	E	1132
Db	1764	---	---	L	O	A	E	L	S	N	E	---	---	1799	
Qy	1133	K	T	E	V	D	E	E	K	---	---	---	---	1185	
Db	1800	K	L	O	E	V	E	G	A	V	A	K	L	S	1842

```

Qy 1186 AENARCKVCPKGEDEDDKLLILDCENKAFHLFCLRPALYEVDPGEWQCPACQAPATARNSR 1241
Db 1843 -----SLKQKDKK-----LKEVLLQVEDERKWAEOYKEQAEGNTK 1878
Qy 1246 GRNYTE--ESASDSEDESDESEEEEEEEEEEDYEAGLRRLRPKKT--IRGKH5VIPP 1301
Db 1879 VKQLKQLEAEABESQCINANRRKLORELDATESNEAGREVNALKSLRRGNEASFVP 1938
Qy 1302 AARSGRR 1308
Db 1939 SRRAGGR 1945

RESULT 29
US-09-764-176-7
; Sequence 7, Application US/09764176
; Patent No. 6809189
; GENERAL INFORMATION:
; APPLICANT: NOTEBORN, Mathieu Hubertus Maria
; APPLICANT: DANEN-VAN OORSCHOT, Astrid Adriana Anna Maria
; APPLICANT: ROHN, Jennifer Leigh
; APPLICANT: WEISS, Bertram
; TITLE OF INVENTION: APOPTIN-ASSOCIATING PROTEIN
; FILE REFERENCE: 4725US
; CURRENT APPLICATION NUMBER: US/09/764,176
; CURRENT FILING DATE: 2001-01-17
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 7
; LENGTH: 1400
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: misc feature
; OTHER INFORMATION: Description of Sequence: Amino acid sequence deduced from
; OTHER INFORMATION: leic acid sequence of AAP-
US-09-764-176-7

Query Match 3.7%; Score 297; DB 4; Length 1400;
Best Local Similarity 19.4%; Pred. No. 1.7e-12;
Matches 308; Conservative 245; Mismatches 551; Indels 484; Gaps 78;

Qy 3 PLIG--RKPPP-----LVNPLP----GEPFFFTPIHTQAFR-----TREEYEARKLY 45
Db 25 PLLDLPELPPELVERLVAAPPDVGNGEVPKELVHLKLMRKIKGKSVTADRMKEYLYIKY 84
Qy 46 SERIWTCKSTGSQLTHKEAWEEQEVABELL-----KKEFPWYKELVLEWHHTASL 99
Db 85 -----LCEQFDNLLKPKNIINEEDATVRLQPIGRDKGLMYWYQ---LDQDDHVRMYI 136
Qy 100 EKLVD---TAWLEIMTKYAVGBECDPEVGKEKMLVKVIKIHPLEKVDEATEKKSDDGAC 156
Db 137 EQDDQDGSSW-----KCIVRNRNELAETLALKAQ-IDPVLKNSSQDQNSR 184
Qy 157 DSPSSKENSQIAQHQKKEIVYVADSLIRTERPPNKELIVRYFIRINALRAGTGENAPVWED 216
Db 185 EFSLEDEET-----KKEBETPKQEEQKSEKMKSEEQPMDL----- 221
Qy 217 FLPHKYDVKLQNEBKIIISNVPADSLIRTERPPNKELIVRYFIRINALRAGTGENAPVWED 276
Db 222 -----ENRSTANVLEETTVKGEKEDEKELVK-----LPVIVKL 254
Qy 277 ELVKYKSLSPKSFDFLLDPVKYMTLNPSTKRNKTGSPDRKPKSKKTDNSSLSPNPKL 336
Db 255 E----KPLPE-----NEEKIKIGESDFKENVKPIKVEVKECRADPKDTKS----- 297
Qy 337 WCHVHLKKSLSGSPLYKXNSKNSK-----PEEHLEENMKOMSPN-----KLHTN 381
Db 298 ----SMEKPAQEPERIEFGCNIKSSHEITEKSTETKLNQDQAKIPLKKEIKLSDD 353
Qy 382 FHIPKKGPI-----PAKKPGKHGDK-----PLKAKG-RSKGILNGQKS-----TGN 420
Db 354 FDSVPKGLCKSVTPNKEFLKDKIKOBEETCKRISTITALGHGKOLVNGEVSDEKRVAPN 413

```



```
QY 236 VPADSLITERPPNKEIVRYFIRHNALRAGTCENAPWVVEDELVKYSLPSKFS---DPL 292
Db 847 -----EAAARQKLQ---LEKVTAAEKIKKLEDEILVMDQNNKLSKERKLL 889
QY 293 LDPYKVMTLNPSTKRKNTGSPDRKPSKSKTDSLSPLNPKLWCHVHLKSLSG--- 348
Db 890 BERISDLTNNLAEBEKA-----KNLTCLKKNKHESMISEL-----EVLKKEEKSROEL 938
QY 349 SPLKYVNSKNSKSPPEHLEEMKQMSPNKLHTNFHPPKGPAPKPGKHSDKPLKAKGRS 408
Db 939 EKLKREKLGSDADPHEQIADLQAIAELKM---QLAKKEBELQAALRLDDEIAQKNA 994
QY 409 -KGI-----LNGKSTGNSKSPK-----GLKTPKTKMKTLLDMAKGTOK 449
Db 995 LKKTIRELGHTSDLOEDLDSEARAARNAKAKQKRDGLGELEALKTELED--TLD-STATQ 1051
QY 450 MTRAPRNSGGT-----PRTSSKPHKHLPPAALHLIAYYKENKDREKRSALSCVIST 502
Db 1052 ELRAKREQEVTVLKALDEETRSHEAQVQEMRQKIAQAVELTEQLEQBFKA-KANLDKN 1110
QY 503 ARLLSSEDRARPEELRSLVQKRYELLEHHKRWASMBEQRKEYLUKKREBELKKLEKA 562
Db 1111 KQTEKEN-ADLAGELRVLQAQKE-VEHKKK-----KLEAQVQELQSKCD-- 1155
QY 563 KERRKEMLERLEKQRYEDQELTKNLPAPFLVDTPGELPNTLPGDVAMVVEFLSCYSG 622
Db 1156 GERARAEIENDVKHLQON-EVESVTG-----MLNEAEGKAIKLAKDVASLSSQLDQTQE 1207
QY 623 LLLPDAQPIITAVSLMEALSADKGGFLVNLRVILLQTLQTLQDAIED----- 674
Db 1208 LLQESTROKLVNSTKLRQLEEBRNS-----LQDQDEMEAKQNL 1248
QY 675 --YSGELGKMLSEIPLTHSVSLVR-LCLRRSDVOESESGSDTDNDKDSAAFE-----DN 726
Db 1249 RHISTNLQSDSKKQLQDFASTVEALBEGKRFQKEIENLTQQVEEKAAYDKLEKTKN 1308
QY 727 EVQDEFE-----LEKLTSEPELTSBKKLQILTLACHRLMTYVQDHH 770
Db 1309 RLQQLDLDLVLDNQRLVNSLEK-KORKFDQLAEEK-----NISKYA----- 1353
QY 771 ETRQMSAEL-WKERLAV-----LKEENDKRAEKOKRKEKMEAKNKENGKENGVLG 820
Db 1354 DERDRAEAREKETKALSARALEAELEKELEERTNMLKAENEDLVSSKDDVGKVVH 1413
QY 821 KTDKRXKRVKFPQVD-----TEADMTISAVKSRL-----LAIQAKEREIOER--- 865
Db 1414 ELEKSKRAL--ETQMEEMKTQLELEDELEQATEDAKRLVNMQALKQGFERDLQARDEQ 1471
QY 866 --EMKVKLQ-----AEEERIRKHAARAKAFQEGIAKAKLVWRTPPI-GTDRNHR 915
Db 1472 NEEKRQLOQLUHEYETELEDERKORALAAAKKLGDLKDLQLEQADSAIKGRE----- 1526
QY 916 YWLFSDVEPGLPIEGWVHDSIDYRNFHCKDHTVSGDEDYC---PRSKANLGNKASNN 972
Db 1527 -----EAIKQLKLAQMKD-----FORELEPARASROEIPATAKENEKAKSLADLMQ 1576
QY 973 TQHGAT-----EVAVETTPPKQONLWFLCDSQKELDELLNCLHPQIGRES 1019
Db 1577 LQEDLAAARARKQADLEKELEAEELASSLSGRNA--LQDEKRRLE-----ARIA 1624
QY 1020 QLKERLEKRYQDIHIIHILARKPNLGLKSCDGNQELNPLFRSDDLIVATRLQKGLGYVE 1079
Db 1625 QLEEELESE-----QGNMEAM-----SDRVKATQ----- 1649
QY 1080 ETSSEFEARVISLEKLDGFCGVIALQASVKKFLQGFMAPKQKRKLOSEDSAKTEEYDE 1139
Db 1650 -----QAEOLSN-----LATERSTAQK-----NESARQOQLERQNKELRSKUHE 1688
QY 1140 EKKMVE-----EAKVASALEBKWTAITAREAQTFMRVLLGLMDLACIKWMSAENA 1189
Db 1699 MEGAVSKFKSTIAALEAKIAEQVEQEAEREQAATK----- 1727
```

```
QY 1190 RCKVCPKKGEDDKLILCDECNKAFHLCFLRPALEYVDPGEWQCPACOPATARRNSGRNY 1249
Db 1728 -----SLKQKDK-----LKBILLQVEDERKMAEQYKEQAKGNARVQL 1767
QY 1250 TE--ESASEDESDDESDEEEEBEEDYEVAGLRLRPKT--IRGKHSVIPPAARS 1305
Db 1768 KRQLEAEESQRIANRRKLQRELDDEATESNEAMGREVNALKSKLRGNETSFVPSRRS 1827
QY 1306 GRR 1308
Db 1828 GGR 1830

RESULT 32
US-09-538-092-1084
; Sequence 1084, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
; APPLICANT: Giot, Loic
; APPLICANT: Mansfield, Traci A.
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538.092
; CURRENT FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CuratPatSeqFormatter Version 0.9
; SEQ ID NO 1084
; LENGTH: 1972
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (0)-(0)
; OTHER INFORMATION: Polypeptide Accession Number P35749
; US-09-538-092-1084

Query Match 3.6%; Score 290.5; DB 4; Length 1972;
Best Local Similarity 19.9%; Pred. No. 8e-12;
Matches 275; Conservative 214; Mismatches 507; Indels 387; Gaps 58;

QY 75 LKBEFFAWYKLVLEMVHNTASLEKLVDTAMLEIMTKY-----AVGEED 121
Db 801 LARKAFARQOQLTAMKVIQRNCAAYLKLNRWQWRFLTKVKPLLVQVTRQEEEMQAKDE 860
QY 122 FEVGKEMLVKVIKHPLEKVDDEATEKKS-----DGACDSPSSDKENSQIAQDHQK 175
Db 861 LQTKEROQKAE-NELKELEQKHSQLTKEKLLQEQLOAQETELYAEAEEMVRLAAKQE 919
QY 176 KETVYKDEGRRESINDRARRSPRKLPTSLKKGKRWAPPKFLPHKYDVKLQNEDKIISN 235
Db 920 LEEILHEARLESEEDRGQ-----LQAEK-----KMAQQLDLEQLEB----- 961
QY 236 VPADSLITERPPNKEIVRYFIRHNALRAGTCENAPWVVEDELVKYSLPSKFS---DPL 292
Db 962 -----EAAARQKLQ---LEKVTAAEKIKKLEDEILVMDQNNKLSKERKLL 1004
QY 293 LDPYKVMTLNPSTKRKNTGSPDRKPSKSKTDSLSPLNPKLWCHVHLKSLSG--- 348
Db 1005 BERISDLTNNLAEBEKA-----KNLTCLKKNKHESMISEL-----EVLKKEEKSROEL 1053
QY 349 SPLKYVNSKNSKSPPEHLEEMKQMSPNKLHTNFHPPKGPAPKPGKHSDKPLKAKGRS 408
Db 1054 EKLKREKLGSDADPHEQIADLQAIAELKM---QLAKKEBELQAALRLDDEIAQKNA 1109
QY 409 -KGI-----LNGKSTGNSKSPK-----GLKTPKTKMKTLLDMAKGTOK 449
Db 1110 LKKTIRELGHTSDLOEDLDSEARAARNAKAKQKRDGLGELEALKTELED--TLD-STATQ 1166
QY 450 MTRAPRNSGGT-----PRTSSKPHKHLPPAALHLIAYYKENKDREKRSALSCVIST 502
```

Db 1167 ELRAKEQEVTVLKKALDEBTRSHQAQVQMRQKHAQVEELTQEQFKRA-KANLDRN 1225
QY 503 ARLSSSEDRARLPEELRSIVQKRYELLEHHKXWASMSBEQRKEYLKKREELKKLKEKA 562
Db 1226 KQTEKEN-ADLAGELRVLGQAQOE-VEHKK- - - - -KLEAQVQELQSCSD- 1270
QY 563 KERREKEMLERLEKQKRYQDBELTGKPLPAFLVDTPEGLPNTLFGDVAMVVFELSCYSG 622
Db 1271 GERARAEALNDKVHKLQK-EVESVTG- - - - -MLNEAGKAIKLAKDVASLSQLQDTQE 1322
QY 623 LLLPDAQYPTITAVSLMEALSADKGGFLYLNRLVILLOTLLQTLQDETAED- - - - - 674
Db 1323 LQOETROKLVNSTKLQLEEBRNS- - - - -LQDQDEMEAKQNL 1363
QY 675 --YGLGKMLSEIPLTLHVSSELVR-LCLRRSDVQESSESGSDTDNDKDSAAFE- - - - -DN 726
Db 1364 RHISTINIQLSDSKKKLQDFASTVEALEBEGKRFQKEIENLTQOYEEKAAAYDKLETKN 1423
QY 727 EVQDEF- - - - -LEKLTSEFFELTSEKQLITLALCHRLMYSVQDHM 770
Db 1424 RLQELDDLVDLNDQRLVSNLEK-KQKFPQLLAEK- - - - -NISSKYA- - - - - 1468
QY 771 ETRQMSAEL-WKERLAV- - - - -LKEENDKRAEKQKREKEMAKNKGKVENGLG 820
Db 1469 DERDRAEAAREKETKALSARALBEALBELEKRTNKLKAEMEDLVSSKDDVGNVH 1528
QY 821 KTRDKKRIKVPPOVD- - - - -TEAEDMISAVKSRL- - - - -LAIQAKKEREIOER- - - - - 865
Db 1529 ELEKSKRAL- - - - -ETQOEMTKQLELEDELQATDACLRLVNMQLKGFQERDLQARDQ 1586
QY 866 --EMVKVLRQ- - - - -AEEERIRKHAARAKAFOEGIAKAKLVMRPTI-GTDNRNHR 915
Db 1587 NEEKRQLQRLQHEYTELEDERKQPALAAAKKLEGLDKOLELQADSIAKRE- - - - - 1641
QY 916 YWLFSEVPLGFIKQGWHDSDYRPNHCKDHTVSGDEYD- - - - -PRSKANLGRNAGN 972
Db 1642 ----EAIQLRKLQAKMD- - - - -FORELEDARASRDEIFATAKENEKKALEADLMQ 1691
QY 973 TQGTAT- - - - -EVAVETTPKQONLWFLCDSQKDELINCLHPQGIRES 1019
Db 1692 LOEDLAARARAKQADLEKEBELAELASLSGRNA--LQDEKRLLE- - - - -ARLA 1739
QY 1020 QLKERLEKRYQDIHSHILARKPNLGLKSCDGNQELLNPLRGLDIEVATRLQKGLGYVE 1079
Db 1740 QLEEELEEB- - - - -QGNMEAM- - - - -SDRVKATQ- - - - - 1764
QY 1080 ETSEPEARVISLEKLDGFCGVIALQASVIKFLQGFMAPKQKRLQSEDSAKTEEYDE 1139
Db 1765 ----QAEQLSNE- - - - -LATERSTAQK- - - - -NESARQQLERQNKELSKLHE 1803
QY 1140 EKKMYE- - - - -EAKVASALEKWKTAIREAQTFSRMHVLLGMLDACIKWDSAEVA 1189
Db 1804 MEGAVKSKFSTIAALEKIAQLSEQVEQEAKEQAATK- - - - - 1842
QY 1190 RCKVCPKGEDDKLILDCBNKAFHLFLCLRPALYVEPDGEMQCPACOPATARNRGRNY 1249
Db 1843 ----SLKQDKK- - - - -LKEILLQVEDERKMAEQVKEQAEKGNARVQL 1882
QY 1250 TE--ESASDESDDESEEEEEEBEYEVAGLRRLPRKT--IRGKHSVIPPAAARS 1305
Db 1883 KQLEEAEBESQRINANRRKQRELDDEATESNEANGREYNALSKLRRGNETSFVPSRRS 1942
QY 1306 GRR 1308
Db 1943 GGR 1945

RESULT 33
US-09-949-016-7111
; Sequence 7111, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:

; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 7111
; LENGTH: 1984
; TYPE: PR1
; ORGANISM: Human
; US-09-949-016-7111

Query Match 3.6%; Score 290.5; DB 4; Length 1984;
Best Local Similarity 19.9%; Pred. No. 8.1e-12;
Matches 275; Conservative 214; Mismatches 507; Indels 387; Gaps 58;
QY 75 LLKEFFPAWYKEL-VLEMVHNTASLEKLVDTAWLEIMTKY- - - - -AVGEECD 121
Db 813 LARKAFKQQQLTAMKVIQRNCAAYLKLNNQWWRLLFTKVKPQLLVTRQBEEMQAKEDE 872
QY 122 FEVGKEMLVKVIKIHLEKVDSEATEKKS- - - - -DGACDPSDSDKENSQIAQDHQK 175
Db 873 LQTKEROQKAB-NELKELEKHSQLTBEKNLQLOQAETELYAEAEEMRVLAAKQOE 931
QY 176 KETVVEDEGGRESINDRARRSPKLPSTLKKGERKWAAPPKPLPHKYDVKLQNEDKI 235
Db 932 LEEILHEARLEEBEDRGQ- - - - -LQAEKK- - - - -KAAQOMLDEEQL- - - - - 973
QY 236 VPADSLIRTERPPNKEIVRYIRHNALRAGTCENAPWVVEDELVKYSLPSKFS- - - - -DPL 292
Db 974 ----EEAARQKLQ- - - - -LEKVTAEAKIKLEDEILVMDQNNKLSKERKL 1016
QY 293 LDPYKMTLNTSTKRNKNTGSDPRKPSKSKTDNSSLSPNLPKLMCHVHLKLSLG- - - - - 348
Db 1017 EERISDLTTLNLAEEBEKA- - - - -KNLTKLKNKHSIMISEL- - - - -EVLKKEERSQBL 1065
QY 349 SPLKVNKSNKSPPEELHEEMKMMSPNKLHTNFI PKKPPAKPKGKHSKPLKAKGRS 408
Db 1066 EKLKRLKLGSDAFHEQIADLQAIKELM- - - - -QLAKKEELQALARLDDDEIAQKNA 1121
QY 409 -KGI- - - - -LNGQKSTGNKSPKK- - - - -GLKTPKTKMQLTLDMAKGTQK 449
Db 1122 LKKIRELGHSIDLQEDLDSERAARNAEKQKRDIGEELEALKTELED- - - - -TLD-STATQ 1178
QY 450 MTRAPRNSGGT- - - - -PRTSSKPKHKLPPAALHLIAYYKKNKDRKRSALSVISKT 502
Db 1179 ELRAKEQEVTVLKKALDEBTRSHQAQVQMRQKHAQVEELTQEQFKRA-KANLDRN 1237
QY 503 ARLSSSEDRARLPEELRSIVQKRYELLEHHKXWASMSBEQRKEYLKKREELKKLKEKA 562
Db 1238 KQTEKEN-ADLAGELRVLGQAQOE-VEHKK- - - - -KLEAQVQELQSCSD- - - - - 1282
QY 563 KERREKEMLERLEKQKRYQDBELTGKPLPAFLVDTPEGLPNTLFGDVAMVVFELSCYSG 622
Db 1283 GERARAEALNDKVHKLQK-EVESVTG- - - - -MLNEAGKAIKLAKDVASLSQLQDTQE 1334
QY 623 LLLPDAQYPTITAVSLMEALSADKGGFLYLNRLVILLOTLLQTLQDETAED- - - - - 674
Db 1335 LQOETROKLVNSTKLQLEEBRNS- - - - -LQDQDEMEAKQNL 1375
QY 675 --YGLGKMLSEIPLTLHVSSELVR-LCLRRSDVQESSESGSDTDNDKDSAAFE- - - - -DN 726
Db 1376 RHISTINIQLSDSKKKLQDFASTVEALEBEGKRFQKEIENLTQOYEEKAAAYDKLETKN 1435
QY 727 EVQDEF- - - - -LEKLTSEFFELTSEKQLITLALCHRLMYSVQDHM 770

Db 1704 LQEDLAAERARQADLEKEELAEELASSLSGRNA--LQDEKRRLE-----ARIA 1751
QY 1020 QLKERLEKRYQDI IHSIHLARKPNLGLKSCDGNQOELLNPLRSLIEVATRLQKGGGLGYVE 1079
Db 1752 QLEEELEEE-----QGNWEAM-----SDRVKATQ-----1776
QY 1080 ETSEFEARVISLEKDFEGECVIALQASVIKFFLOGFMAPKQKRRKLQSEDSAKTBEVD 1139
Db 1777 -----QAEQLSNE-----LATERSTAQ-----NESARQQLERQNKELRSLKHE 1815
QY 1140 EKKQVTE-----EAKVASALEKWKTAIREAQTFPSRMHVLLGLMDACIKWDSAEWA 1189
Db 1816 MEGAVKSKFKSTIAALEAKIAQLEEQVEQAEKQAATK-----1854
QY 1190 RCKVCPKGGEDDKLILCDENKAFHLFCLRPALYVDPDGEWQCPACOPATARRNSGRNY 1249
Db 1855 -----SLKQKDK-----LKEILLQVEDERKVAEQYKEQAEKGNARKVQL 1894
QY 1250 TE--ESASEDSEDESEEEEEEEEDYEVAGLRPRKT--IRGKHVSIPPAARS 1305
Db 1895 KRQLEAEESQRINANRRKLQRELDATESNEAMGREVNALKSLRRGNETSFPVPSRS 1954
QY 1306 GRR 1308
Db 1955 GGR 1957

RESULT 35
US-09-949-016-7113
; Sequence 7113, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7113
; LENGTH: 1984
; TYPE: PRF
; ORGANISM: Human
US-09-949-016-7113

Query Match 3.6%; Score 290.5; DB 4; Length 1984;
Best Local similarity 19.9%; Pred. No. 8.1e-12;
Matches 275; Conservative 214; Mismatches 507; Indels 387; Gaps 58;

QY 75 LLKEEFPANVEKL-VLEMVHNTASLEKLVDTAWLEIMTKY-----AVGECD 121
Db 813 LARKAFAGQQQLTAMKVIFQRCAAYLKURNQMWRLFTKVKPLQVTRQEEEMQAKDE 872
QY 122 FEVGKEKMLKVKIVKIHPLKDEEATEKKS-----DGACDSPSDKENSQIAQDHQK 175
Db 873 LQTKERQQKAB-NELKELEQHSQLTTEEKNLLQEQLAETELAYAEAEEMRVRLAAKQOE 931
QY 176 KETVVKEDGREGSINDRARRPRKLPSTSLKIGERKWAAPPKLPKHYDVKLQNEDKIISN 235
Db 932 LEEILHEMEARLEEEEDRGQQ-----LQAEK-----KMAQQMLDEEQLEE-----973
QY 236 VPADSLIRTERPPNKEIVRYFIRHNALRAGTCENAPWVVEDELVKKYSIPSKFS---DPL 292
Db 974 -----BEAARQKQLO-----LEKVTAEAKIKKLEDEILVNDQNNKLSKERKLL 1016

QY 293 LDPYKYMTPNTPKRNKTSPPDRKPSKSKTDNSSLSPLNPKLWCHVHLKKSLSG-----348
Db 1017 EERISDLTNTLNAEEKA-----KNLTCLKNKHESMISEL-----EVALKKEEKSQOEL 1065
QY 349 SPLKVKNKSNKSPBEHLEEMKMSPNKLTNFIHPKGGPPAKPKGKISDKPLKAKGRS 408
Db 1066 EKLKRLKLEGDASDFHEQTADLQAOIABLKN-----QLAKKEEELQAALARLDBEIAQKNN 1121
QY 409 -KGI-----LNGOKSTGNSKSPK-----GLKTPKTKMKQMTLLDMDAGTK 449
Db 1122 LKIKLELEGHISDLOEDLDSERAARNKAKQKRDGEELEALKTELED--TLD-STATQQ 1178
QY 450 MTRAPRNSGGT-----PRTSSKPHKHLPPAALHLIAYYKKNKDRKRSALSCVISKT 502
Db 1179 ELRAKREQEVTVLKALDEBTRSHQAQVQEMRQKIAQAVEELTEQLEQFKRA-KANLDKN 1237
QY 503 ABLLSSEDPARLPBELRSIVOKRYELLEHHKKWASMESEORKEYLKKKEBELKKLUKEKA 562
Db 1238 KQTLKEN-ADLAGELRVLGQAKQB--VEHKCK-----KLEAQVQELQSKCSD--1282
QY 563 KERRREKEMLERLEKOKRYEDQELTGKNLPAFRLVDTPEGLPNTLFGDVAMVVEFLSCYSG 622
Db 1283 GERARAEALNDKVHKLQN-EVESVTG-----MLNEAEKKAIKLAKDVASLSQLODQOE 1334
QY 623 LLLPDAQYPITAVSLMEALSADKGGFLYLNRLVILLQTLQTLQDEIAED-----674
Db 1335 LLQEEETROKLVNSTKLQLEEBRNS-----LQDQLEDEMEAKQNLE 1375
QY 675 --YGEIGHMKLSIPLTLHSVSELVR-LCLRRSDVQEESESGSDTDNKKDAAFE-----DN 726
Db 1376 RHISTLNTQLSDSKKQLODFASTVEALEEGKRFQKEIENLTQOYEEKAAAYDKLEKTKN 1435
QY 727 EVQDEF-----LEKLTSEFPPELTSEEKLQILTALCHRLMILTSYVDHM 770
Db 1436 RLQOEELDLVDLQNRQVLSNLEK-KQKFFOLLAEK-----NISKYA-----1480
QY 771 ETRQMSABL-WKERLAV-----LKEENDKKRAEKQKREKMEAKNKGKVENGLG 820
Db 1481 DERDRAEAEREKETKALSARALBEALEKEBELRTNKMKAEMEDLVSSKDDVGKNVH 1540
QY 821 KTDKRRKRVKPEPOVD-----TEADMTSAVKSRL-----LAIQAKKEBEIQR---865
Db 1541 ELEKSKRAL--ETOMEEMKTLQEELEDELQATEDAKLRLEVNMQALKGQFERDLOARDEQ 1598
QY 866 --EMKVKLERO-----AEEERIRKHKAASAKAFOEGIAKAKLVNRRTP1-GTDRNHNK 915
Db 1599 NEEKRRQORQLHEYTELEDERKORALAAAKKLEGLDKOLELOQADSAIKGRE-----1653
QY 916 YWLFSDVEVPGLFIEKGWVHDSIDYRFNHHCKDHTVSGDEDYC---PRSKKANLGNKASVN 972
Db 1654 -----EAIKQLRKLQAKMD-----FORELEDAARASRDEIFATAKENEKKAKSLEADLMQ 1703
QY 973 TOHGAT-----EVAVETTPKQOQNIWFLCDSOKELDELINCLHPQGIRES 1019
Db 1704 LQEDLAAERARQADLEKEELAEELASSLSGRNA--LQDEKRRLE-----ARIA 1751
QY 1020 QLKERLEKRYQDI IHSIHLARKPNLGLKSCDGNQOELLNPLRSLIEVATRLQKGGGLGYVE 1079
Db 1752 QLEEELEEE-----QGNWEAM-----SDRVKATQ-----1776
QY 1080 ETSEFEARVISLEKDFEGECVIALQASVIKFFLOGFMAPKQKRRKLQSEDSAKTBEVD 1139
Db 1777 -----QAEQLSNE-----LATERSTAQ-----NESARQQLERQNKELRSLKHE 1815
QY 1140 EKKQVTE-----EAKVASALEKWKTAIREAQTFPSRMHVLLGLMDACIKWDSAEWA 1189
Db 1816 MEGAVKSKFKSTIAALEAKIAQLEEQVEQAEKQAATK-----1854
QY 1190 RCKVCPKGGEDDKLILCDENKAFHLFCLRPALYVDPDGEWQCPACOPATARRNSGRNY 1249
Db 1855 -----SLKQKDK-----LKEILLQVEDERKVAEQYKEQAEKGNARKVQL 1894
QY 1250 TE--ESASEDSEDESEEEEEEEEDYEVAGLRPRKT--IRGKHVSIPPAARS 1305

```
Db 1895 KQLEAEESORINARRKLORELDATESNEAMGREVNALKSLRRGNETSFPVPSRRS 1954
Qy 1306 GRR 1308
Db 1955 GGR 1957

RESULT 36
US-09-595-684B-31
; Sequence 31, Application US/09595684B
; Patent No. 6544766
; GENERAL INFORMATION:
; APPLICANT: Beraud, Christophe
; APPLICANT: Ohashi, Cara
; APPLICANT: Sakowicz, Roman
; APPLICANT: Vaisberg, Eugeni
; APPLICANT: Wood, Kenneth
; APPLICANT: Yu, Ming
; TITLE OF INVENTION: Human kinesins and methods of producing
; TITLE OF INVENTION: and purifying human kinesins
; FILE REFERENCE: cytop036
; CURRENT APPLICATION NUMBER: US/09/595,684B
; CURRENT FILING DATE: 2002-06-24
; PRIOR APPLICATION NUMBER: 09/295,612
; PRIOR FILING DATE: 2000-04-20
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 31
; LENGTH: 2662
; TYPE: PRT
; ORGANISM: Human
US-09-595-684B-31

Query Match 3.5%; Score 280; DB 4; Length 2662;
Best Local Similarity 19.2%; Pred. No. 7,1e-11;
Matches 333; Conservative 307; Mismatches 608; Indels 490; Gaps 79;

Qy 39 EARLERYSERIWTCKSTGSSQITHKAEWEEQVAAELKKEEPAPWYKLVLEMHVHTAS 98
Db 788 ESRVQGLLEEIGKTKDLDATTQSNYSKTDQEFQNFKTLHMDPEQKY-KNVLE--ENERM 843
Qy 99 LEKLVDTAWLEMTKYVAGEECDFVGEKMLKVKI-VKIHLEKVDDEATEKSGDACC 157
Db 844 NOEIVN-----LSKEA--OKFDSLG--ALKTELSYKTOELQKTEVQERLNE---- 888
Qy 158 SPSSDKENSQIAQDHQKKTETVVKEDGRRRESINDRARRSPRLPT-SLAKGERKWAPPK 216
Db 889 -----MEQLKEQLENRDSPLQTVREKTLITEKLQOTLEEVKTLTQEKDDLKQLQES 940
Qy 217 FUPHKYDVKLQNEDKIISNPADSLIR-----TERPNKEIVRYFIRHNALRAGT 266
Db 941 LQIERDQLKSDIHDVTVMNIDTQEQURNALSLKHQETINTLKSISEVSRNLHMEENT 1000
Qy 267 GENAPWVVEDELVKYSPLSPKESFLLDPYKVTWLNPSKRNKTSQDRKPSKSKTDN- 325
Db 1001 GE-----TKDEFQQQKVGIDKKQD--LEAKNTQTITADVKNONEITEQQRKPSLSIOEKNE 1053
Qy 326 --SSLSPLNPKLWCHVHLKSLGSPKLVKNKSNKSPPEEHLEEMKMSPNKLTNPFH 383
Db 1054 LQOMLESVTAKEQLKTDLKENIE--MTIENQBELRLGLDELKQKQEIIVAGKQKHA--- 1107
Qy 384 IPKGGPPAKPKGHSKDKPLAKGRSGKILNGQKSTGN-----SKSPK-----GLTKPTK 433
Db 1108 IKKEGELSRTDLRAVEEKLKESQOLQKQQLLVQEMSEMOMKINETENLKN-EL 1166
Qy 434 YMKQWTL-----LDMAK-----GTQKMTAPNSGGTGR----- 462
Db 1167 KNEKLTLEHMETERLELAQKNENVEEVKSIYKERVKLQKSPETERDHLRGVIREIE 1226
Qy 463 -TSSKPHKLPALHLYAYKENKDRED--KRSALSCVISKITARLLSSED-----RARLP 515
Db 1227 ATGLQTKELKTAHILH-----KEHQETIDELRS-----VSEKTAQIINTQDLEKSHTKLQ 1278
```

```
Qy 516 EELSLVQKRYELLEHKKRWASMSG-----EQRKEYLKKKRE 552
Db 1279 EEI-PVLHEEQELLPNVKKVSETQETWNELELLTEQSTTKDSTTIARIEMERLRUNERFQ 1337
Qy 553 ELKKKLKEKAKERRE-----KEMLERLEKQKRYEDQELTGNLPAF 593
Db 1338 ESQBEIKSLTKERDNLKTIKEALEVKGHDOLKEHIRETLAKIQESQSKQESQSLNKE---- 1393
Qy 594 RLVDTPPEGLPNTLFGDVAMVVEFLSCYSGLLLPDQYDITAVSL-----MEALSA 643
Db 1394 -----KDNETTIVSEMEQFKPKDSALL-----RIEIMLGLSKRLQESHDENKSVAK 1441
Qy 644 DKGGLYLNRLVILLQTLQLLQ-----DEIADYCELGKMLSE 684
Db 1442 EKDDLQRLQEVLESQDLKENIKEIVAKHLEETBEELKVAHCCLKBQSETINELRVNLS 1501
Qy 685 -----IPLTLHSVSELVRLCLRRSDVQSESGSDTDDNKDSAAFEN----- 726
Db 1502 KETESTITQKQLEAINDKL-----QNKIQEIYEKEEQLNIKQISEVQENVNELKQFKEHR 1556
Qy 727 EVQDEFLEKLETSFEFFELT-----SEKQLQILTALCHRLMTYSVQDHMETRQMSAELW 781
Db 1557 KAKDSALQSIIE-SKMELTNRLOESQEBIQIMIKEEM--KRVQEAQOIERDOLKENT 1612
Qy 782 KERLAVLKEENDK-----KRAEKQKQKEM--EAKNKENGKVGKGLGKTD 823
Db 1613 KEIVAKMKESQEKYQFLKMTAVNETQKMCIEIHLKEQFQTKLUNLENIETEN----- 1666
Qy 824 RKKRIVKPEPOVTEAEDMISAVKSR---RLLAQAKKEREIQEREMVKLERQAEERIR 880
Db 1667 -----IRLTOILHENLEEMRSVTKERDDLRSVEETLKVVERDOLKENLETTITDLEKQEE 1721
Qy 881 RKHKAAPAFQEGIAKAK-LVMERT-----PIGTDNRNRYWLFSDVPGVPIEGKWVHD 935
Db 1722 LKIVHMLKHEQETIDKLGIVSEKTNISNNQDLHSNDALKAQD---LKIQE----- 1773
Qy 936 SIDIRFNH-HCKDHTVSGDEDYCPRSKCANLGKNASMTQHGCTATEVAVEITTPQGGNL 994
Db 1774 --ELRIAHMLKHEQETIDKLGIVSEKTKLGNMQKOLENSNA----- 1815
Qy 995 WFLCDSQKELDELNCLH-----POGIRESQLK-----ERLEKRYQDIHSHLAKPNLG 1045
Db 1816 -----KLQEKIQELKANEHQLITLKQDVNETQKQVSEMEQLKKQIKD--QSLTSLK--LE 1866
Qy 1046 LKSCDGNQELNFRSDLIIEVATRL-QKGGIGYVEETSEFEARVLSLKLKDFGSCVIAL 1104
Db 1867 IENLNLAQE---LHENLEEMKSVMKERDNLRRVEETLKLER-----DQLKESLOETKAR 1917
Qy 1105 QASVIKKFLOGFMAPKQKR---KQSDSASAKTEVEDEBKKWBEAKVASALEKWKTAIR 1161
Db 1918 DLEIQOELKTAARMLSKEHETVQKUREKISEKTIQISDIQKOLDKSK-----DELOKKIQ 1972
Qy 1162 EAQTFSRMHVLLGLMDACIKWMDMSAENAR---CKVCPKKGEDDKLILCECNKAFHLFCL 1218
Db 1973 ELQK-KEQLLR-----VKEDVNMSHKKINEMEQKKQFEPNYLCKEMDN-----FQL 2020
Qy 1219 RPALYVEPDGWCQCPACOPATARRNSRGRNTYESASEDSEDESEDESEDESEDESEDE 1278
Db 2021 TKKLHESLE-ETRIVA-----KERDELRRIKESLKMRED 2053
Qy 1279 YEVAGLRPRKTRGKHSVIPPAARSGRPRCKK-----PHSTRRSQPKAPVDDAEVD 1332
Db 2054 QFIATLRMIARD-RQNHQV-----KPEKRLSDGQOQHLMESLREKC-----SRIK 2098
Qy 1333 ELVLQTKRESS-----RRQSLEQKCEETLHKI---VKYRFSWPPFPVTRDEADY 1380
Db 2099 ELL---KYSSEMDHDHYECLNRLSLDLK-ELIEFHIMKKLYVLSY-----VTKIEEQH 2149
Qy 1381 YDVITHPMDF-----QTVNKCSCGSYR-----SVQEFITDMQK 1414
Db 2150 ECINKFEMDFIDEVEKQKELLIKIQHLOQDCDVPSPRELRLDLKLNQMDLHIEILKD---- 2206
```


QY 1483 EDRRGREABIOEWLOQTSYLSAKINSKOHNCFMMLVNTQFCMALTDT 1529
DB 3007 VSDS--WSEIR--DDEAFARVKEEQKIFGLVNDQRSQGTTPDT 3049

RESULT 40

US-08-056-200-94
; Sequence 94, Application US/08056200
; Patent No. 5616500
; GENERAL INFORMATION:
; APPLICANT: Steinert, Peter M.
; APPLICANT: Lee, Seung-Chul
; APPLICANT: Kim, In-Gyu
; APPLICANT: Chung, Soo-Il
; APPLICANT: Park, Sang-Chul
; TITLE OF INVENTION: Trichohyalin and Transglutaminase-3 and
; TITLE OF INVENTION: Methods of Using Same
; NUMBER OF SEQUENCES: 117
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 620 Newport Center Drive, Sixteenth Floor
; CITY: Newport Beach
; STATE: CA
; COUNTRY: U.S.A.
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/056,200
; FILING DATE: 30-APR-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Fedrick, Michael F.
; REGISTRATION NUMBER: 36,799
; REFERENCE/DOCKET NUMBER: NIH054.001A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (714) 760-0404
; TELEFAX: (714) 760-9502
; INFORMATION FOR SEQ ID NO: 94:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1898 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-056-200-94

Query Match 3.4%; Score 272; DB 1; Length 1898;
Best Local Similarity 18.9%; Pred. No. 1.7e-10;
Matches 203; Conservative 188; Mismatches 417; Indels 264; Gaps 37;

QY 483 KENKREDKRSALSCVSKTARLLSSDRARLPEELRSILVQKRYELLEHKKRWASMSBEQ 542
DB 288 RERQEEQQOQLR--REQLRKQEEERREQEEERREQEEERREQEEERREQEEERREQEE 345
QY 543 RKE-YLKKRKEELKKK---LKKAKERREKEML-----ERLEKQKRYEDQELTGKNLPAP 593
DB 346 RREQLRRQEEERREQEEERREQEEERREQEEERREQEEERREQEEERREQEEERREQEE 395
QY 594 RLVDPEGLPNTLFGDVAMVVEFLSCYSGLLLPDAQYPITAVSLMEALSADKGGFLYNR 653
DB 396 -----LRR 398
QY 654 VLVLILQTL---QTLQDEIAEDYCELGKMLSEIPLTLHVSSELVR-LCLRRSDVQES 709
DB 399 EQQLRREQLRREQLRREQLRREQLRREQLRREQLRREQLRREQLRREQLRREQLRREQLR 453
QY 710 EGSDDDNKDSAAFEDNEVDQEFLEKLETSSEFELTSEKQLILTALCHILMTYSVQDH 769
DB 454 EQHEQERREQLRREQLRREQLRREQLRREQLRREQLRREQLRREQLRREQLRREQLRREQL 499

QY 770 METROQSAELWKERLAVLKEENDKKRAEKQKREMEAKNKENGKVENGLGKTRDKKRI- 828
DB 500 EER-----RERWLKEEE---ERREQERREQLRREQEEERREQLRREQEEERLQ 548
QY 829 --VKPEQVDTAEADMISAVKSRLLAIQAKKERIEIQEREMKVKLERQAEERIRKHAA 886
DB 549 QRLRSEQQLRREQEEERL-----EQLLKREERKLEERREQLRREQEEERDQLKRE-- 601
QY 887 AEKAFQEGIAKAKLVMRPTPIGTDNRNHRNYLFDVPGFLFIEKGWVHSDYDRENNHCK 946
DB 602 -EERQQLKREQ-----EERLEQR--LKREEVERLEQEE-----R 635
QY 947 DHTVSGDEDYCPRSKKAN--LGKNASMTQHGTAATEVAVETTPKQGNLWFLCDSQKEL 1004
DB 636 DERLKKEE---PEEERHELLKSEQEERREHQLRREQEEERREQLRKE-----EEREL 687
QY 1005 DELNCLHPQIGRESQL-----KERLEKRY-----QDIHIIHARKPNLGL 1046
DB 688 EQLRKREHEERREQLAEERQAEARERIKSRIPKWQWQLESEADARQSKVLEAPQAGR 747
QY 1047 KSCDGNQELLNPLRSDL-----LEVATRLQKGLGVVEETSFEA-----RVISLEK 1093
DB 748 ABAFQEEKRRRESELQWQEEERAHQQQEEERDRDFTWQWQAEKSEKRGQRUSARPP 807
QY 1094 LKDFGECVIALQASVKKFLQGFMAPKOKR--RKLQSDSDAKTEEDVEDEKMWVEAKVAS 1151
DB 808 LREQE--RQLRABERQEQEORFLPEEBEKEQRGQRERKEQLQFLEEEQLORRERAQ 865
QY 1152 ALEKWKTAIRBAQTFSRMVLGLMDACIKWDMGAENARKVCPKKGDDDKLILCENK 1211
DB 866 QLQEEEDGLQEDQERRRQE--QRRDQKRWQLEBEERKRR----- 903
QY 1212 AFHLFCLRPALYE-----VPDGEWQCPACQAPATARNRGRNYTESASEDSEDDDES 1263
DB 904 --HTLYAKPALQEQLRKEQLLQEEELQREERKRRRQEQERYREEQLQEEELQ 961
QY 1264 DEEEEEEEDYEVAGLRPRKTIRGHVSIPPAAARSGRPFKPKHSTRSQPKA 1323
DB 962 REERKRRRQERQY-----RKDKLQKEEQL-----LGEPEKRRRQEREKRYE 1009
QY 1324 PPVDDAEVDELVLQTKRSSRQSLQEQ--KCEILHKIVKYRFPSPFPVTRDEADYY 1381
DB 1010 EELQEEELQREERKRRQEWERYRKQDELQEE-----BEQLLREERK-- 1057
QY 1382 DVIHTPMDFTVONKSCGVSYSVQEFLLTMKQVFTNAEVYNCRSHVLSWVKTEQCIV 1441
DB 1058 -----RRLQERER--QYRBEELQEEBEQLL--GEERTRRRQLERQYRKEEELQ 1104
QY 1442 VLLHKHLPCHPVYRRKKKFPDLRLAEDGDSPEPAVGQSRDDEDRSRRAEIQ 1493
DB 1105 QEEEQLLREPEKRRRQERERQCRREEELQEEELQEEELQREERKRRRQELERQ 1156

RESULT 41

US-08-800-644-94
; Sequence 94, Application US/08800644
; Patent No. 5958752
; GENERAL INFORMATION:
; APPLICANT: Steinert, Peter M.
; APPLICANT: Lee, Seung-Chul
; APPLICANT: Kim, In-Gyu
; APPLICANT: Chung, Soo-Il
; APPLICANT: Park, Sang-Chul
; TITLE OF INVENTION: Trichohyalin and Transglutaminase-3 and
; TITLE OF INVENTION: Methods of Using Same
; NUMBER OF SEQUENCES: 117
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 620 Newport Center Drive, Sixteenth Floor
; CITY: Newport Beach
; STATE: CA
; COUNTRY: U.S.A.
; ZIP: 92660

COMPUTER READABLE FORM:	
MEDIUM TYPE: Floppy disk	
COMPUTER: IBM PC compatible	
OPERATING SYSTEM: PC-DOS/MS-DOS	
SOFTWARE: PatentIn Release #1.0, Version #1.25	
CURRENT APPLICATION DATA:	
APPLICATION NUMBER: US/08/800,644	
FILING DATE: 14-FEB-1997	
CLASSIFICATION: 424	
PRIOR APPLICATION DATA:	
APPLICATION NUMBER: US 08/056,200	
FILING DATE: 30-APR-1993	
ATTORNEY/AGENT INFORMATION:	
NAME: Fredrick, Michael F.	
REGISTRATION NUMBER: 36,799	
REFERENCE/DOCKET NUMBER: NIH054.001A	
TELECOMMUNICATION INFORMATION:	
TELEPHONE: (714) 760-0404	
TELEFAX: (714) 760-9502	
INFORMATION FOR SEQ ID NO: 94:	
SEQUENCE CHARACTERISTICS:	
LENGTH: 1898 amino acids	
TYPE: amino acid	
TOPOLOGY: linear	
MOLECULE TYPE: protein	
US-08-800-644-94	
Query Match	3.4%; Score 272; DB 2; Length 1898;
Best Local Similarity	18.9%; Pred. No. 1.7e-10;
Matches	203; Conservative 188; Mismatches 417; Indels 264; Gaps 37;
QY	483 KENKREDKRSALSCVISTKARLLSSDDRRLPEELRSVLQKRYELLEHKKRWASSEEQ 542
DB	288 RREQEEOQQLR--REQLRRKQEEERREQEERREQEERREQEERREQEERREQEER 345
QY	543 RKE-YLKKREELKKK--LKEKAKERREKEML-----ERLEKQRYEDELTKGNLPAP 593
DB	346 RREQLRREQEERREQEERREQEERREQEERREQEERREQEERREQEERREQEER 395
QY	594 RLVDTPGLPNTLFGDVAMVVFSLSCVSGLLLPDAQYPTITAVSLMEALSADKGGFLYLN 653
DB	396 -----LRR 398
QY	654 VLVLILQTL---QTLQDIABDYGBELGKMLSEIPLTHSVSELVR-LCLRRSDVQEE 709
DB	399 EQQLRREQLRREQLRREQLRREQLRREQLRREQLRREQLRREQLRREQLRREQLR 453
QY	710 EGSDDDNKOSAFEDNEVDQEFLEKLETSFELTSEELKQLILTLCHRLMTYVSQDH 769
DB	454 EQKHEQRRQLRREQEERRLKREETERHE--QERRKQQLK-----RDQ 499
QY	770 METRQMSAELMERLAVLKEENDKKRAEKQKEMAKNKENGKVENGLTKTRKKRI- 828
DB	500 EER-----RERWLKLEE---ERREQEERREQLRREQEERREQLRQEEERLQ 548
QY	829 --VKFEPQVDEADMISAVKSRLLAIQAKKREIQEREMKVKLRQAEERIRKHAA 886
DB	549 QRLRSEQLRREQEERL-----EQLLKREKLEQLERREQLRQEEERDQLKRE-- 601
QY	887 AEKAFEGIAKALVMRTPIGTRDNRHNYWLFSDVEPGLFIBKGWVHDSIDYRPNHCK 946
DB	602 -ERRQQLRREQLRREQEERREQLRREQEERREQLRREQEERREQLRREQEER 635
QY	947 DHTVSGDEDYCPRSKCAN--LGNASMNTOGTATEVAVTTTPKQGNLWFLCDSQKEL 1004
DB	636 DERLKREE---PEERREHLLKSEQEERREQLRREQEERREQLRREQEERREQLR 687
QY	1005 DELLNCLHPGIRESQL-----KERLEKRY-----QDIHSHLARKNGL 1046
DB	688 EQLKREHEERREQLAEQEQAERIKSRIPKQWQLESDARQSKVLEAFAQR 747
QY	1047 KSCDGNQELNLRSDI-----IEVATRLKGGGLGYEETSEPA-----RVISLEK 1093

```
QY 654 VLVILLQTL---QTLQDEIAEDYGLGKMLSEIPLTLHSVSELVR-LCLRRSDVQRES 709
Db 399 EQLRRQQLRREQQLRREQQLRREQQLRREQQLRREQQLRREQQLRREQQLRREQQLR 453
QY 710 EGSDDDDNDKSAFEDNEVDQBFLEKLTSEFFELTSEBKQLITLALCHRLIMTYSDVQH 769
Db 454 EQKHEQEREQRLKREQEERRDLKREETERHE--QERRKQOLK-----RDQ 499
QY 770 METROOMSABLWKLAVLKEENDKKRAQKQKMEAKNKNKGVENGKGTDRKKRI- 828
Db 500 EBER-----RERWLKEE---ERREQERREQERREQERREQERREQERREQERREQ 548
QY 829 --VKPEQVDTAEADWISAVKSRLLAIOAKKERIQREREMVKLERQAEERIRKXAA 886
Db 549 QLRREEQQLRREQERL-----EQLLKREERKLEQERREQERREQERREQERREQER 601
QY 887 AKAFAQEGIAKALVMRTPICTDRNNHRYLFSDEVPGLFTEKGWVHDSIDYRFNHCK 946
Db 602 -ERRQQLKREQ-----ERLEQR--LKREVERLEQER-----R 635
QY 947 DHTVSGDEDYCPRSKAN--LGNASMTQGTATEVAVETTPPKQGNLWFLCDQKEL 1004
Db 636 DERLKREE---PEERRHLLKSEQERHRLRREQERREQERREQERREQERREQER 687
QY 1005 DELLNCLHPQGTRESQ-----KERLEKEY-----QDILHSLARKPNLGL 1046
Db 688 EQLRKREHEERREQELAEEOQARERIKSRIPKQWOLESEADAROSKVLLLEPAQGR 747
QY 1047 KSCDGNQELLNFRSDL-----LEVATRLQGGGLGYVEETSEPEA-----RVISLEK 1093
Db 748 AEAPOQEERKRESELOQWEEERAHQOEEQERDFTWQOAEKSEKRGQRQLSARPP 807
QY 1094 LKDFGCVIALQAVIKPLOGFMAPKQR--RKLQSEDSAKTEVEDEKKNVBAKVAS 1151
Db 808 LFAEQRE--RLAABERQOERQFLPEEBEKEQGRQREKEKELQLEEEELQRRERAQ 865
QY 1152 ALEKWKTAIREAQTSRHHVLLGMLDACIKWDSAEANARCKVCPKKGDDKLILCDECNK 1211
Db 866 QOEEEDGLQOQERRQO---ORRQKRWOLEBERKRR-----903
QY 1212 APHLFCLPALYE-----VPDGEWCPACQATARRNSRGRNVTSEASESDEDDDES 1263
Db 904 --HTLYAKPALQELRKESQQLQOEEELQREERKRRQOERQYREEEQLQOEEQL 961
QY 1264 DEEBEEREBEEDYEVAGLRLRPKTIIRGHKSVIPPAARSGRRPGKXPHSTRSQPKA 1323
Db 962 REERKRRQERQY-----RKDKLQKEEQ-----LGEPEKRRRQERREKRYRE 1009
QY 1324 PPVDDAEVDLQTKSRSSRQSLQO--KCEILLHKIVKYRFPFPVTRDEADYY 1381
Db 1010 EELQOEEELQREERKRRQERQYREERKRRQOERQYREERKRRQOERQYREERKRR 1057
QY 1382 DVITHPMDFTQVQNCSCGSYRSVQEFLLTDMQVFTNAEVNCRGSHVLSMVKTEQCLV 1441
Db 1058 -----RLQERER--QYREELQOEEQLL--GEERTRRQERQYREERQYREELQ 1104
QY 1442 VLLHKLPHGVYRKRKFPDRLAEDGDSPEAVGQSRDDBRRSRAEIQ 1493
Db 1105 QEEELQREERKRRQERQYREERQYREERQYREERQYREERQYREERQYREERQY 1156
```

RESULT 43

```
US-09-949-016-7659
; Sequence 7659, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
```

```
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7659
; LENGTH: 2753
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-7659
```

```
Query Match 3.4%; Score 270; DB 4; Length 2753;
Best Local Similarity 19.4%; Pred. No. 4e-10;
Matches 327; Conservative 247; Mismatches 606; Indels 508; Gaps 80;
```

```
QY 19 BEPPFTIPIHTQEAFTREEYEARELYSE--RIWTC-KSTGSSQLTHKEAMEEEOEVAEL 75
Db 1337 EEP-----GEPFEIVERVEDLEKNEILRSGTCTRDESSVQSRSERGVLVEEWIV 1389
QY 76 LKSEFPWAYEKLVLMEVHHHTASLEKLVDTAWLEIMTYAVGEECDFFVGKEMKLVKIV 135
Db 1390 SDEETEARQKAPLE-----ITEY-----PC-VEVRIDKEIKGKVE 1424
QY 136 K-----IHPLEKVDDEATEKSDGACDS-----PSSDKE-----NSSQIA 170
Db 1425 KDSTGLVNYLTDLNTCVLPKEQLQTVQDKAGKCEALAVGRSSEKEGKIDPPDETOST 1484
QY 171 ODHOKETVVKEDGESRESINDRARRSPKLPSTLKKGERKWAPKPLPHKYVDVLQWED 230
Db 1485 QKQHPSLGINKPV--RRKLAKEKQKEEGLOASAEALK-----KGSSEE 1529
QY 231 KI-----ISNVPADSLIRTERPPN--KEIVRYPIRHNALRAGTGENAPWVVEDEL 278
Db 1530 SLGEDPGLAPEPLPTVKATSPLEETPIGSIKDKVKALQKR-----VEDEQ 1575
QY 279 VKKYSLPKSFDFLLDPYK-----YMLNLPSTK--RKTGSPDRKPKSKKTDNSLSLS- 330
Db 1576 KGRSKLPTRVKGKEDVPKTTTHRPHAPASPSLKERHAPGSPKPTERHSTLSSAKTER 1635
QY 331 --PLNPKLWCHVHLKSLGSPKVNKSKNSKEEHLKEMKMSPNKLTNHPH- 385
Db 1636 HPPVSPSKTEKHPVSPSAKTERHSPASSSKTEKH-----SPVSPS--TKTERHSPVSS 1689
QY 386 ---KKGPPAKPKGKHSKDK--PLKAKGRSKGLNGOKSTGNSKSPKGLKTPKTKMKQMTL 440
Db 1690 TKTERHPPVPSGK--TDKRPVPSPSGRTE-----KHPP-----1721
QY 441 LDMAGTQKMTAPRPNSSGCTPR-----TSSKPHKHLPPAALHLIAYYKENKDRKRAL 495
Db 1722 VSPQTEKRLPVSP--SGRTDKHPVSTAGTKHLPVSP-----SGKTEKQPPV 1769
QY 496 SCVTSKTARLLASSEDARLPEELSLVQKRYELLEHKKRWASMSSEORKEYLKKREELK 555
Db 1770 S-PTSCTERI-----EETMSV-----RELML--KAFQSGQDPKSKHTGLTFEHSK 1812
QY 556 KKLKAKERERKE---MLERLEKQRYEDQELTGKNLPAFLVDTPEGLPNTLFG----- 608
Db 1813 KQPEKGVRYEKEKGPILTQREAKTENQTIKRGQRLPVTGTAESKRGVRSVSSIGVKE 1872
QY 609 DVAMVVEFLSCVSGLLLPDAQYPTAVSLMALSGDGFLYLNRLVILLQTLLOTLQ 668
Db 1873 DAAGGKERVLSHK---IPE--PVQSVPEEESHRESE-----VPK 1906
QY 669 DEIAEDYGLGKMLSEIPLTLHSVSELVRLCLRRSD-----VOESEGSDTD-----DN 717
Db 1907 EKMADEQDMQLQSPDRKSTDFSEVIKQLENDKTYQQFRLSEETKAQLHLDQVITS 1966
QY 718 KDSAAFEDNEVDQBFLEKLTSEFFELTSEBKQL-----ILTALCHRLIMTYSDVQHMETRQ 774
Db 1967 PFNTTFLDYMKDEFLPALSLQSGALDGSSESLKNEGVAAGSPCGSL-----MEGTP 2017
```



```
QY 775 QMSAE--LWKEPLAVLENDKKRA-EKQKRKEMAKNKENGVENGLTKDRKKRIYKF 831
Db 2018 QISSBESYKHEGLAETPSPESLSPKSPKSGGETKESKTKE-----TTTETRSEKE 2072
QY 832 EQPVDTEADMTSAVKSRLLAIQAKKEREIQEREMKVKLERQAEERIRKHAAAEKAF 891
Db 2073 HP-----TKYDITGGSEERATVTE-----DSETSTESFQKEATLGSFKDTSPPKQ 2118
QY 892 QEGIAKALVM--RRTPIGTDRNHNRYWLFSDVPGLFIEKGWVHDSIDYRFNHCKDHTV 950
Db 2119 DDCGTSCSVALAKETPTGLTEE-----AACDEGQRTFGSS--AHKT--QTDSEVQEST 2168
QY 951 SDEBDCPRSKKANIGKASMMTQCHTATEVAVEITTPKQGNLWFLCDSQKEL-----DE 1006
Db 2169 TSDE-----TKALPLPEASVKTDTGTSKPGQVIRSP--QGLELALPGRDSEVL SAVAD 2221
QY 1007 LNLCLH-----POGIRESQLKERLEKRYQDIIHSHLARKPNLG--- 1045
Db 2222 SLAVSHKUSLEASPVLNDSNHSKTPDSLEPSLK---ESPGRDSLESSPVEPKMAGIFP 2278
QY 1046 ----LKSCDGNQELNFIKSLIEVATRIQKGLGYVEBTSFEARVISLEKLDGFGCV 1101
Db 2279 SHFPLPAVAKTELL-----TEVASVRSLLRPDPGSAEDD-----SLE----- 2317
QY 1102 IALQASVIKFKLOGFMAP--KQKRKLQSEDSAKT-----SEVDBEK----- 1141
Db 2318 ---QTSLMSSGKSPSLPTPSESEVSYEVTKPTDVTTPKPAVTHECAEDDSENGEKK 2374
QY 1142 -----KMYEAAKVASALEK-----WKTATREAOFTGRMHVLLGMLDACIKWMS 1185
Db 2375 RPTPEEMFKWTKIMFDELOEAKQKRDYKKEPKQESS-----SS 2418
QY 1186 AENARKVC---PK---KGEDDK---LILCDECNKAFHLCIRPALYVDPG----- 1228
Db 2419 DPDAKDVDPKHTGSGEDESVPVLVTSRKYVSSSESEPELAQLKKGADSGLLPE 2478
QY 1229 ---EQWCPACQATARNR-----GRNYTEESASESDEDESEDESESESESE 1275
Db 2479 PVIRVQPSPLPSSMDNSPEEVQFPVVSQYTFK--MNETOBEPGKSEBEKSESHL 2537
QY 1276 BDEYVAGLRPRKTIRKHSVIPPAAARSGRRPGKKPHSTRSQKAPPVDDADEVLA 1335
Db 2538 AEDQLAREL-----HALLKY---WLERDQKHATDNLVECLTKINRMD---I 2600
QY 1336 LQTKRSRRQSLQKCEBILKIVKRYFSWPR-----EPVTRDAEDYDYVI 1384
Db 2558 IENPNSLODQS-----HALLKY---WLERDQKHATDNLVECLTKINRMD---I 2600
QY 1385 THPMDPOT---VQNKSCGYSRSVQBEFLT--DMKQVFTNAEVYVNCRSHVLSCHVTEQCLV 1441
Db 2601 VHLMETNTEPLOERIS--HSYAEIEQITLIDHSEGFVSLQBELCTAQH-----KQBEQA 2653
QY 1442 VLLKHHLPGHPVVRKR-----KKFPDRLADEGDSBEPAV--GQSRDEDRRSREAEIQ 1493
Db 2654 VSKESETCDHPPIVSEEDISVGYSTFQDGVPTKTEGSSATALLFPQTHKEQVQDFSGRMQ 2713
QY 1494 EWLQDTSL 1501
Db 2714 DLPESSSL 2721
```

RESULT 45

```
US-09-538-092-1154
; Sequence 1154, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
; APPLICANT: Glot, Loic
; APPLICANT: Mansfield, Traci A.
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538,092
; CURRENT FILING DATE: 2000-03-29
```

```
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CuraPatSeqFormatter Version 0.9
; SEQ ID NO 1154
; LENGTH: 3210
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (0)---(0)
; OTHER INFORMATION: Polypeptide Accession Number P49454
US-09-538-092-1154

Query Match 3.4%; Score 270; DB 4; Length 3210;
Best Local Similarity 18.6%; Pred. No. 4.9e-10;
Matches 312; Conservative 293; Mismatches 574; Indels 498; Gaps 78;

QY 52 CKSTGSSQLTHKE-----AWEEQOEVAELLKEFPAPWYKLVLE--MVHHTASLEKL 102
Db 1624 CSSLOEENLTRKETPSAPAKGVEELESCEVYRQSLKELEKESGQIMKNKEIQELEOL 1683
QY 103 VDT-----AWLEIMTKYAGEECDFEVGKEK-----MLKVKIVIHPL 141
Db 1684 ISSERQELDLRKQVLSNEQWQKLTSTVLEMESKLAEEKQTQLSLELEVARLQ--LQ 1742
QY 142 KYDEEA-----TEKSGDGCADSPSSDKENSQIAQDHQKQKTVVVKEDGGRRESIN--- 191
Db 1743 GDLSSRLGIDTIEDAQGRNESCDISKEHTSETTERTPKHDVHQICDKDAQDNLNLDI 1802
QY 192 DRARRSPKRLPTSLKKGR-----KWAPPKFLPHKYDVQLQNEDKLIISN-----VPA 238
Db 1803 EKITETGAVKTPGEGSGQSPDTNYEPPG-----EDKTQSGSECSISELSFSGPNALVPM 1856
QY 239 DSLIRTPPKKEIVRYFIRN---ALRAGTGENAP--WVVEDELVKVYSLPSKF---S 289
Db 1857 DFL-----GNQEDI-----HNLQVRKETSNEENRLLHVIHEDRDKVESLLENKELDS 1905
QY 290 DFLDLPYKYMT-----LNPSTRKKTGSPDRKPSKSKTDNSSL-----SSPLN 333
Db 1906 KLHLQEVQLMTKIEACIELEKTVGELKKENSDDLSEK--LEYFSCDHQELLQRVETSEGLN 1963
QY 334 PKLVCHV--HLKKSUSGSLPKVONS-----KNKSPEEH-----LEEMKKN 372
Db 1964 SDLEMHADKSSREDIGDNVAKVNSWKEKFLDVENELSRIRSEKASIEHEALYLEADLEV 2023
QY 373 MSPNKLHTNFHPKKGPPAKPGKHSKPLAKGRSGILNGQKSTGNSKSPKGLKTPK 432
Db 2024 VOTEKL-----CLEKDNENKQKVVIVCLEBELSVVTSERNQLRGELDTMSKKT--TALDQLS 2077
QY 433 TKMKQMTLLDMAKGTQKMTAPRNSGGTPTSTSKPHKLPPAALHLIAIYYKENKDRDKR 492
Db 2078 EKKKEK-----TQEL-----ESHQ-----SECLHICV--AAEAVKCKT 2109
QY 493 SALSCVISKTRALLSSEDRARLPPEELRSLVQKRYEL-----LEHKRWASMSSEQRKEYLK 548
Db 2110 ELLQTLSSDVSELL--KDKTHLOEKLSLEKDSQALSITKCELENQIAQLNKE--KELIV 2165
QY 549 KKEBELKKKLKE-----K 561
Db 2166 KESESQARLESSEYKLVNSKALEAALVEKGEFALRSLSSQTEFVHQURRGIEKURVRIE 2225
QY 562 AKERREKEMLERLEKQKRYEQELTGKQLPAPFLVDTPTPEGLPNTLFGDVAMVVEFLSCVS 621
Db 2226 ADEKQKQLHAEKLEKERENDS-----LKDVENLEREL-----QMSSENG 2266
QY 622 GLLLPDAQVPTTAVSLMEALSADKGGFVLYANRVLLIQTLLQTLQDEIADYSELGKMK 681
Db 2267 ELVILDAENSKAENVETLTKQIEEMARSLKVPFLDLVTLRSEKENLTK-QIQEKQGO---- 2321
QY 682 LSEIPLTLHVSVELVRLCLRRSDVQ--ESESEGS-----D 713
```


Qy	239	DSLIRTERPPNKELVIRFIHNN	---	ALRAGTGENAP	---	WVVEDELVKYISLPKPF	---	S	289				
Db	1857	DFL	----	GNQEDI	----	HNQLQRVKYETSNNELRL	LLHHVIEDRDRKVESILLNMKELDS	1905					
Qy	290	DFLLDPVKYMT	----	LNPSTKRKNTGSPDRKPSKSKTDNSL	---	---	SSPLN	333					
Db	1906	KLHLOEVQMLTKI	EACIELESKI	VELKXKNSDISEK	---	LEYFSCDHQELLQRVETSEGLN	1963						
Qy	334	PKLWCHV	---	HLKXSLSGSPLKRVKNS	----	---	KNSKSPPEH	---	LEENMKM 372				
Db	1964	SDLEMHADKSSREDIGD	NVAKVNDWKERFLD	VENELSRIRSEKASIEHAELYEADLEV	2023								
Qy	373	MSPNKHLTNFHI	PKKGPAPKPKGHSOKPL	KAKGRSKGIINGKOSTGNSKSPKKGLKTPK	432								
Db	2024	VQTEKL	----	CLEKDNENKOKVIVCLBEELS	VVTSERNQJRGELDTMSKKT	---	TALDQLS	2077					
Qy	433	TKWKOMTLDMWGTOKW	TRAPNSGGT	PTRTSSKPKHLP	PPAALHLIAYYKENKORDKR	492							
Db	2078	EKKMEK	----	TQEL	---	ESHO	---	SECHLCIQV	---	AAEAUVKEKT 2109			
Qy	493	SALSCVISTARLLSS	DRARLPPELRSLVQKRYEL	----	---	LEHKCRWASMSBQRKEYLK	548						
Db	2110	ELLQTLSSDVSELL	---	KDKTHLEKQSLSEKSOALS	LTKELENQIAQLNKE	---	KELLV	2165					
Qy	549	KDBEELKKUKP	----	----	----	----	----	----	K 561				
Db	2166	KESSELQARLS	ESDYELKNSV	KALEAALVEKGBFALR	SLSTQB	EVHQLRRGIEKLRVRIE	2225						
Qy	562	AKERREKEMLERLEK	OKRYEDQELT	TKNLPAPFLVDTPEGL	PNTLFGDVAMVVEFLSCYS	621							
Db	2226	ADSKQJHIAEKUKER	ENDS	----	---	LKOKVENLEREL	----	OMSEENQ	2266				
Qy	622	GLLLPQAQYPI	TAVSLMEALS	SADKGGPLYLNRLVILLO	TLLOTLQDEIAEDYGBELGMK	681							
Db	2267	ELVILDAENSKA	EVETLKTQI	EBWARS	LKTFELDLVTLRSEKENLTK	---	QIEQKQGO	----	2321				
Qy	682	LSEIPTLTHSVSEL	NRLCLRRSDVQ	---	ERESEGS	----	----	----	D 713				
Db	2322	LSELDKLLSFKSL	ILBE	---	KEQABIQIKESATVEMLQNL	QELNEAVALCGDQBEIMKA	2380						
Qy	714	TDDNKDSAAEDNE	VQDEFLEKLETS	EFFELTSEKLIQ	LITALCHRLIMTYYS	VQDIMEYR	773						
Db	2381	TEQSLPPIIEEH	QHLRNS	---	IEKLRA	----	RELADEKKQ	----	JC	---	VLQOLKSESH	----	2425
Qy	774	QOMSAELWKERL	AVLKEENDKCAE	KOKREMAKKNKENGK	VENGLKTRDKRKRI	VKVFEP	833						
Db	2426	---	HADLLKGRVENLER	ELEIART	---	NOEHAALAEANSK	---	GEVET	----	----	----	2464	
Qy	834	QVDTAEADMT	SAVKRRLLAI	QAKKEBIEI	QIEREMKVKLER	----	QAEE	877					
Db	2465	---	LKAKTEGTMQSLRG	LELDVVVTIRSEKENLT	NELQKEQERISLE	LEINSSFENILOEKEQ	2523						
Qy	878	ERRR	---	KHKAAB	---	KAQEGITA	----	KAK	----	---	LWRRPTI	---	907
Db	2524	EKVOMKEKSTAM	EMLMQTLQ	KELNERNVAALHNDQ	EACAKENQLS	QVCELELEKQALLQ	2583						
Qy	908	GTDRNHRNYL	FSDPEVPLFIE	KGWVHDSIDYR	FHHKDHDTYS	GDGEDYCPRSKCANLKG	967						
Db	2584	GLDEAKNNYIV	LOSSV	----	---	KGLIQEVEDGQKLEK	KDDEEISRLKNQOI	QOQEQLVSK	2636				
Qy	968	NASNMTQHG	PATVAVETTT	PKQGNLWFLCDS	QKELDELLNCLHP	PGTIR	RESOLKLERLEK	1027					
Db	2637	LSQVEGEHQLWKE	----	---	QNL	---	ELRNLTVLELEQIK	QIVLQSKNASLQDTLEVLQS	2684				
Qy	1028	RYODI	THSIHLARKPNL	KSCDGNQELN	FLR	---	SDLI	EVATRLQKGLGYVEETSEFE	1085				
Db	2685	SYKOLENELELTK	----	---	MDKMSFVEKVNKMTAKET	LELOREHMAQKTAELQ	2732						
Qy	1086	ARVISLEKLD	GFBCVIALQAS	VIKKFLQGFMAP	KOKRRKLQSEDS	AKTEEVD	---	BEKKM	1143				
Db	2733	EE	---	LSGEKVR	LAGE	----	---	LQLLLEBI	KSKQOLKSLTJL	ENSELKSLOMHIKQOV	2782		
Qy	1144	VEEAKVASALEK	WKMTAIREAQ	TFSRMHVLLGLMD	DACIKWDM	SAENARCKVCPKKGDDKL	1203						

RESIT.T 48

RESOL 48
US-09-418-710-67

US-95-418-710-07
: Sequence 67. Application US/09418710

; Patent No. 6596482

GENERAL INFORMATION:

APPLICANT: Jones, Michael H.

; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR

; FILE REFERENCE: 06501-042001

; CURRENT APPLICATION NUMBER: US/09/418,710

10 ; CURRENT FILING DATE: 1999-10-15

; PRIOR APPLICATION NUMBER: PCT/JP98/01783

; PRIOR FILING DATE: 1998-04-17

;
PRIOR APPLICATION NUMBER

; PRIOR FILING DATE: 1997-10-24

; PRIORITY APPLICATION NUMBER: JP 9/116570

; PRIOR FILING DATE: 1997-04-18

; NUMBER OF SEQ ID NOS: 7.

```

; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 67

```

; SEQ ID NO 67
: LENGTH: 59

```

; LENGTH: 59
; TYPE: PBT

```

; TYPE: PRT
; OPACNTSM:

ORGANISM: HOMO SAPIENS
IIS-09-418-710-67

US-09-418-10-61

Query Match 3.3%: Score 265: DB 4: Length 59:

Query match 3.3%; score 263; DB 47
Best Local Similarity 90.8%; Pred. No. 3.7e-12;

DDBE Local Similarity: 50.00, Local NC: 5.70 12;
 Matches: 59; Conservative: 0; Mismatches: 0; Indels: 6; Gaps: 2;

[illegible]

QY 630 YPITAVSLMEALSADKGGFLYNRVLVILLQTLLQTLLQDEIAEDYGELGMKLSEIPLTL 68

Db 1 YPI TAVSL

RESULT 49

US-09-839-479-66

; Sequence 66, Application US/09839479

; Patent No. 6727222

GENERAL INFORMATION:

APPLICANT: Jones, Michael H.

; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR

FILE REFERENCE: 06501-042002

Search completed: March 9, 2005, 15:01:14
Job time : 99.9379 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 9, 2005, 13:06:37 ; Search time 142.925 Seconds
(without alignments)
5336.298 Million cell updates/sec

Title: US-10-702-148-21
Perfect score: 10138
Sequence: 1 MGQYKTSSTSGGGRKCNQEQ.....AGHNRYKFEKKWDTFKVS 1972

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 65 summaries

- Database : A_Geneseq_16Dec04:*
- 1: Geneseqp1980s:*
 - 2: Geneseqp1990s:*
 - 3: Geneseqp2000s:*
 - 4: Geneseqp2001s:*
 - 5: Geneseqp2002s:*
 - 6: Geneseqp2003as:*
 - 7: Geneseqp2003bs:*
 - 8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	10138	100.0	1972	6	ABR64241 Angiogen
2	10138	100.0	1972	8	ADP54420 Human PRO
3	10131	99.9	1972	2	AAW81171 Human BAZ
4	4153	41.0	796	4	AAM93402 Human pol
5	4153	41.0	796	8	ADL30971 Human pro
6	3559	35.1	708	4	ABR93073 Human pro
7	2903	28.6	589	8	ADR09338 Human pro
8	2513.5	24.8	1878	4	AAW40239 Human pro
9	2512	24.8	1873	7	ABM85417 Human pro
10	2498	24.6	1727	4	ABR95554 Human pro
11	2493.5	24.6	1878	2	AAW81170 Human BAZ
12	2492	24.6	1589	4	AAW42025 Human pro
13	2300	22.7	1586	7	ABM85416 Human pro
14	1244	12.3	244	4	AAW63263 Human bre
15	1078.5	10.6	524	8	ADR09529 Human pro
16	1021	10.1	194	4	AAW14092 Peptide #
17	1021	10.1	194	4	ABR33037 Peptide #
18	1021	10.1	194	4	AAW26499 Peptide #
19	1021	10.1	194	4	ABR27866 Human pep
20	1021	10.1	194	4	ABR18510 Protein #
21	1021	10.1	194	4	AAW66222 Human bon
22	1021	10.1	194	4	AAW53836 Human bra
23	1021	10.1	194	4	ABG47890 Human liv
24	1021	10.1	194	4	AAW01830 Peptide #
25	1021	10.1	194	5	ABG35872 Human pep

26	817	8.1	160	4	AAU16212 Human nov
27	817	8.1	160	6	ABU55281 Human nov
28	762.5	7.5	555	3	ABR21050 Human nuc
29	727.5	7.2	1540	2	AAW07734 Human HTL
30	726.5	7.2	1674	2	AAW81169 Human BAZ
31	726.5	7.2	1674	7	ADP69139 Human MP5
32	726.5	7.2	1674	8	ADP12578 Protein e
33	638.5	6.3	280	6	ABU70931 Human adi
34	582	5.7	134	4	AAU16625 Human nov
35	582	5.7	134	6	ABU55694 Human nov
36	558	5.5	779	4	ABR95562 Human pro
37	536	5.3	1483	8	ADQ20910 Human sof
38	533	5.3	1483	7	ADP69140 Human MP5
39	533	5.3	1531	2	AAW81173 Human BAZ
40	532	5.2	1527	2	AAW81172 Human BAZ
41	491	4.8	97	4	ABR37253 Peptide #
42	491	4.8	97	4	AAW30718 Peptide #
43	491	4.8	97	4	ABR32005 Peptide #
44	491	4.8	97	4	ABR22549 Protein #
45	491	4.8	97	4	AAW70385 Human bon
46	491	4.8	97	4	AAW57960 Human bra
47	491	4.8	97	5	ABG40029 Human pep
48	481	4.7	1476	4	ABR58706 Drosophil
49	463	4.6	91	4	AAW17059 Peptide #
50	463	4.6	91	4	ABR36058 Peptide #
51	463	4.6	91	4	AAW29552 Peptide #
52	463	4.6	91	4	ABR30877 Peptide #
53	463	4.6	91	4	ABR21456 Protein #
54	463	4.6	91	4	ABR69228 Human bon
55	463	4.6	91	4	AAW56843 Human bra
56	463	4.6	91	4	ABG50904 Human liv
57	463	4.6	91	4	AAW04764 Peptide #
58	463	4.6	91	5	ABR38836 Human pep
59	410	4.0	1715	4	ABR58089 Drosophil
60	401	4.0	475	4	ABR93746 Human pro
61	393	3.9	2759	6	AAO16418 Human nuc
62	390.5	3.9	2897	4	ABR58514 Drosophil
63	390	3.8	2781	3	AAW57453 Human tra
64	390	3.8	2781	8	ADF42724 Human BPT
65	390	3.8	2781	8	ADQ18653 Human sof

ALIGNMENTS

RESULT 1	ABR64241	standard; protein; 1972 AA.
ID	ABR64241	standard; protein; 1972 AA.
XX	ABR64241;	
AC	ABR64241;	
XX	ABR64241;	
DT	15-OCT-2003	(first entry)
XX	15-OCT-2003	(first entry)
DE	Angiogenesis protein BNO150.	
KW	Cytostatic; antirheumatic; antiarthritic; antidiabetic; ophthalmological;	
KW	antiproliferative; antiarteriosclerotic; cardiact; vasotrophic; angiogenesis;	
KW	gene therapy; vasculature; cancer; rheumatoid arthritis; psoriasis;	
KW	diabetic retinopathy; cardiovascular disease; atherosclerosis;	
KW	ischemic limb disease; coronary artery disease.	
XX	Homo sapiens.	
OS	WO2003027285-A1.	
XX	WO2003027285-A1.	
PN	03-APR-2003.	
PD	03-APR-2003.	
XX	19-SEP-2002; 2002WO-AU001282.	
XX	19-SEP-2002; 2002WO-AU001282.	
PR	27-SEP-2001; 2001AU-00007973.	
PR	27-SEP-2001; 2001AU-00007974.	
PR	11-OCT-2001; 2001AU-00008210.	
PR	29-OCT-2001; 2001AU-00008532.	

PR	13-NOV-2001; 2001AU-00008838.	Db	421	EEDEDEDDDESDDSDSQSESDNSSESDTEGSEEDDDDDKQDQDESDDTEGEKTSMLNKTT	480
PR	28-AUG-2002; 2002AU-00951032.	Qy	481	SSVKSPSMSLTGHTSTPRNLHIKAPGSAPALCSESQSPAFICGTSSTLTSSPSHSGTSKR	540
XX	(BTON-) BIONOMICS LTD.	Db	481	SSVKSPSMSLTGHTSTPRNLHIKAPGSAPALCSESQSPAFICGTSSTLTSSPSHSGTSKR	540
XX	Gamble JR, Hahn CN, Vadas MA;	Qy	541	RRVTDERELRIPLEYGWQRETRIRNFGRLQGEVAYYAPCGKKLRQYPEVIKYLNRNGIM	600
XX	WPI; 2003-354655/33.	Db	541	RRVTDERELRIPLEYGWQRETRIRNFGRLQGEVAYYAPCGKKLRQYPEVIKYLNRNGIM	600
XX	N-P5DB; ACF34516.	Qy	601	DISRDNFSAKIRVGDYFYEARDGPQBMQWCLLKEDVIPIRAMEGRRRPPNPDRQRA	660
PT	New angiogenic genes and polypeptides, useful for diagnosing,	Db	601	DISRDNFSAKIRVGDYFYEARDGPQBMQWCLLKEDVIPIRAMEGRRRPPNPDRQRA	660
PT	prognosticating or treating an angiogenesis-related disorder, e.g.	Qy	661	REESMRPRKGRPPNVGNABFLDNADAKLRKQAOEIAEQAAQIKLLRKLOEQEARVA	720
PT	cancer, rheumatoid arthritis, diabetic retinopathy, psoriasis or	Db	661	REESMRPRKGRPPNVGNABFLDNADAKLRKQAOEIAEQAAQIKLLRKLOEQEARVA	720
XX	cardiovascular diseases.	Qy	721	KEAKKQQAIIAAAEKRRKQEQIKIMKQOEIKRIQQIRMEKELRAQQIILEAKKKKEEAA	780
PS	Claim 15; SEQ ID NO 176; 90pp; English.	Db	721	KEAKKQQAIIAAAEKRRKQEQIKIMKQOEIKRIQQIRMEKELRAQQIILEAKKKKEEAA	780
XX	The invention relates to the isolation of novel genes (ACF34446-ACF34559)	Qy	781	NAKLEAEKRIKEKEMRRQOAVLLKHQERERRRRQHMMKAMARKKAAEKERLQEKRD	840
CC	encoding proteins (ABR64180-ABR64281) involved in the process of	Db	781	NAKLEAEKRIKEKEMRRQOAVLLKHQERERRRRQHMMKAMARKKAAEKERLQEKRD	840
CC	angiogenesis. The nucleic acid molecules are useful in identifying and/or	Qy	841	EKRLNKERKLEORLELEMAKELKKPNEDCLADQKPLPELPRIPIGLVSGSTFSDCLMV	900
CC	obtaining full-length human genes involved in an angiogenic process. The	Db	841	EKRLNKERKLEORLELEMAKELKKPNEDCLADQKPLPELPRIPIGLVSGSTFSDCLMV	900
CC	nucleic acid molecule, polypeptides or complexes encoded, cells or	Qy	901	VOFLRNFKVLGPDVNIIDVNLVQLQGLNIGDSMGEVQDILLVLLLSAAVCDPGLITGY	960
CC	genetically modified non-human animals derived from these are useful for	Db	901	VOFLRNFKVLGPDVNIIDVNLVQLQGLNIGDSMGEVQDILLVLLLSAAVCDPGLITGY	960
CC	the screening of candidate pharmaceutical compounds used in treating	Qy	961	KAKTALGEHLNVLGNVRDNYSEILQIFMEAHCGQTELTESLTKTKAFAQHTPAQKASVLAF	1020
CC	angiogenesis-related disorders. They are also useful for diagnosing,	Db	961	KAKTALGEHLNVLGNVRDNYSEILQIFMEAHCGQTELTESLTKTKAFAQHTPAQKASVLAF	1020
CC	prognosticating or treating an angiogenesis-related disorder, which	Qy	1021	LINELACKSVVSEIDKNIDYMSNLRDKVWVEGKLRLRIIHAKKTGKRTSGGIDIGE	1080
CC	involves uncontrolled or enhanced angiogenesis or is a disorder in which	Db	1021	LINELACKSVVSEIDKNIDYMSNLRDKVWVEGKLRLRIIHAKKTGKRTSGGIDIGE	1080
CC	a decreased vasculature is of benefit (e.g. cancer, rheumatoid arthritis,	Qy	1081	EQHPLGTPTPKRRRRKGGSDYDDDDDDSDQDDEDEDEKQKKTDTICDEDED	1140
CC	diabetic retinopathy, psoriasis or cardiovascular diseases such as	Db	1081	EQHPLGTPTPKRRRRKGGSDYDDDDDDSDQDDEDEDEKQKKTDTICDEDED	1140
CC	atherosclerosis), or involves inappropriately arrested or decreased	Qy	1141	EGDQAASVELEKQIEKLSKQQSQYRRKLPDASHLSLVSVMFGPDYRRRRYWLPRCGGIF	1200
CC	angiogenesis or is a disorder in which an expanding vasculature is of	Db	1141	EGDQAASVELEKQIEKLSKQQSQYRRKLPDASHLSLVSVMFGPDYRRRRYWLPRCGGIF	1200
CC	benefit (e.g. ischemic limb disease or coronary artery disease). The	Qy	1201	VEGMSGEGLEETAKEREKLIKAAESVQIKEEMPETSGDSLNCNSTDHCEOKEDLKEKDN	1260
CC	modulator of expression or activity of the polypeptide encoded by the	Db	1201	VEGMSGEGLEETAKEREKLIKAAESVQIKEEMPETSGDSLNCNSTDHCEOKEDLKEKDN	1260
CC	nucleic acid sequence is useful for manufacturing a medicament for the	Qy	1261	NLFLOKPGSPKSLKLELVAKMPPSEVMTPKPNAGANGCTLSYQNSGKHSLSGVSQSTAT	1320
CC	treatment of an angiogenesis-related disorder. This sequence corresponds	Db	1261	NLFLOKPGSPKSLKLELVAKMPPSEVMTPKPNAGANGCTLSYQNSGKHSLSGVSQSTAT	1320
XX	to one of the novel angiogenic protein	Qy	1321	QSNVEKADSNLNFMTSGSGFKFYSPLPNDQLLKTLLTEKNRQWFSLLPRTPCDDTSLTHA	1380
SQ	Sequence 1972 AA;	Db	1321	QSNVEKADSNLNFMTSGSGFKFYSPLPNDQLLKTLLTEKNRQWFSLLPRTPCDDTSLTHA	1380
	Query Match 100.0%; Score 10138; DB 6; Length 1972;	Qy	1381	DMSTASLVTQSQPPSPSKSPFTTAPLGSQAQNPVGLNPPFALSPLOVKGVSMMGLQFCGW	1440
	Best Local Similarity 100.0%; Pred. No. 0;	Db	1381	DMSTASLVTQSQPPSPSKSPFTTAPLGSQAQNPVGLNPPFALSPLOVKGVSMMGLQFCGW	1440
	Matches 1972; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	Qy	1441	PTGVVTSNIPFTLTVSPSLGSLGLISEGNNGNSFLTSTNVASSKSPSPQNEKATSAQPAAV	1500
Qy	1 MGQTKSTSSGGNRKCNQEQSKNQLDARVDKI KDKPKRKAMESSNSDSCTSGSDTS	Db	1441	PTGVVTSNIPFTLTVSPSLGSLGLISEGNNGNSFLTSTNVASSKSPSPQNEKATSAQPAAV	1500
Db	1 MGQTKSTSSGGNRKCNQEQSKNQLDARVDKI KDKPKRKAMESSNSDSCTSGSDTS	Qy	1501	EVAKPVDVFPSPKPIPEEMQFGWRIIDPEDLKALLKVLHLRGIREKALQKQIKHLDYIT	1560
Qy	61 SEGISSDSDDLDEDEEEDQSIESEDDSDSSEAEQHKSNQVLLHGISDPKADGQKA	Db	1501	EVAKPVDVFPSPKPIPEEMQFGWRIIDPEDLKALLKVLHLRGIREKALQKQIKHLDYIT	1560
Db	61 SEGISSDSDDLDEDEEEDQSIESEDDSDSSEAEQHKSNQVLLHGISDPKADGQKA	Qy			
Qy	121 TEKAQEKRIHOPLAFPSQTHSFQSQKQPVLSQQLPFTFQSSQAKESVNKHTSVIQ	Db			
Db	121 TEKAQEKRIHOPLAFPSQTHSFQSQKQPVLSQQLPFTFQSSQAKESVNKHTSVIQ	Qy			
Qy	181 STGLVSNVKPLSLVNOAKETMYKLI VPSPDVLKAGNKNTSEESLLTSELRSKREYKQ	Db			
Db	181 STGLVSNVKPLSLVNOAKETMYKLI VPSPDVLKAGNKNTSEESLLTSELRSKREYKQ	Qy			
Qy	241 APPQQLKKQESSKSLKKVIAALSNPKATSSSPAHPKOTLENNHNPNFTNALLGNHPNG	Db			
Db	241 APPQQLKKQESSKSLKKVIAALSNPKATSSSPAHPKOTLENNHNPNFTNALLGNHPNG	Qy			
Qy	301 VIQSVIOBAPLALTTKTKQSKINENIAAASSTPFSFPVNLSTSGRRTPGNQTPVMPAS	Db			
Db	301 VIQSVIOBAPLALTTKTKQSKINENIAAASSTPFSFPVNLSTSGRRTPGNQTPVMPAS	Qy			
Qy	361 PILHSQGEKAVSNVNPVKTOHSHHPAKSLVEQFRGTDSDIPSSKQSEDSNEDEEEDDE	Db			
Db	361 PILHSQGEKAVSNVNPVKTOHSHHPAKSLVEQFRGTDSDIPSSKQSEDSNEDEEEDDE	Qy			
Qy	421 EEDEDEDDDESDDSDSQSESDNSSESDTEGSEEDDDDDKQDQDESDDTEGEKTSMLNKTT	Db			

[illegible]

RESULT 2	
ADP54420	
ID	ADP54420 standard; protein; 1972 AA.
XX	
XX	
ADP54420;	
XX	
18-NOV-2004	(first entry)
DT	
XX	
XX	Human PRO protein sequence SEQ ID NO:396.
XX	
XX	human; PRO; immune related disease; inflammatory immune response;
KW	immune response stimulation; anti-allergic; antianaemic; antiarthritic;
KW	antisthmatic; antidiabetic; antiinflammatory; antipsoriatic;
KW	antirheumatic; antithyroid; CNS; dermatological; gastrointestinal;
KW	haemostatic; hepatotropic; immunostimulant; immunosuppressive; muscular;
KW	nephrotropic; neuroprotective; osteopathic; respiratory; vasotropic;
KW	virucide; gene therapy.
XX	
OS	Homo sapiens.
XX	
PN	WO2004039956-A2.
XX	
PD	13-MAY-2004.
XX	
XX	
PF	28-OCT-2003; 2003WO-US034381.
XX	
PR	29-OCT-2002; 2002US-0422472P.
XX	
PA	(GETH) GENENTECH INC.
XX	
PI	Aggarwal S, Clark H, Gurney AL, Schoenfeld J, Williams PM;
PI	Wood WI, Wu TD;
XX	
DR	WPI; 2004-376182/35.
DR	N-PSDB; ADP54419.
XX	
PT	New PRO polynucleotides and polypeptides, useful in diagnosing
PT	and treating an immune related disease, e.g. systemic lupus
PT	erythematosus, rheumatoid arthritis, diabetes mellitus or asthma and in
PT	stimulating an immune response.
XX	
XX	Claim 1; SEQ ID NO 396; 3009pp; English.
PS	
XX	

The present invention describes an isolated PRO nucleic acid (1). Also described: (1) a vector comprising (1); (2) a host cell comprising the vector of (1); (3) a process for producing a PRO polypeptides; (4) an isolated PRO polypeptide; (5) a chimeric molecule comprising the polypeptide of (4) fused to a heterologous amino acid sequence; (6) an antibody which specifically binds to a polypeptide of (4); (7) a composition of matter comprising a polypeptide of (4), an agonist or antagonist of the polypeptide or an antibody that binds to the polypeptide in combination with a carrier; (8) an article of manufacture comprising a container, a label on the container and a composition of matter of (7); (9) a method of treating an immune related disease in a mammal; (10) a method for determining the presence of a PRO polypeptide in a sample suspected of having the polypeptide; (11) a method of diagnosing an immune related disease or an inflammatory immune response in mammal; (12) a method of identifying a compound that inhibits or mimics the activity of or expression of a gene encoding a PRO polypeptide; and (13) a method of stimulating the immune response in a mammal. The PRO sequences have antiallergic, antianaemic, antiarthritic, antiasthmatic, antidiabetic, antiinflammatory, antipsoriatic, antirheumatic, antithyroid, CNS, dermatological, gastrointestinal, haemostatic, hepatotropic, immunostimulant, immunosuppressive, muscular, nephrotropic, neuroprotective, osteopathic, respiratory, vasotropic and virucide activities, and can be used in gene therapy. The nucleic acid (1) and the encoded polypeptides, compositions, kits and methods are useful in diagnosing and treating an immune related disease and in stimulating an immune response. The present sequence represents a human PRO protein from the present invention.

Sequence 1972 AA:

Query Match	100.0%;	Score 10138;	DB 8;	Length 1972;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 1972;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	MGQTKTSSGGNRKCNQEQSKNQPLDARVDKIKDKPKRKAMSSSNSSDSDSGTSSDTS	60	
Db	1	MGQTKTSSGGNRKCNQEQSKNQPLDARVDKIKDKPKRKAMSSSNSSDSDSGTSSDTS	60	
Qy	61	SEGISSSSDSDLEDEDEEEDQSIIESEDDSDSESAQHKSNNQVLLHGISDPKADGQKA	120	
Db	61	SEGISSSSDSDLEDEDEEEDQSIIESEDDSDSESAQHKSNNQVLLHGISDPKADGQKA	120	
Qy	121	TEKAQEKRIHOPIPLAFESQTHFSQOQKOPQVLSQOLPEIFOSSQAKESVKNKHTSVIQ	180	
Db	121	TEKAQEKRIHOPIPLAFESQTHFSQOQKOPQVLSQOLPEIFOSSQAKESVKNKHTSVIQ	180	
Qy	181	STGLVSNVKPLSLVNOAKKETYMKLVPSPDVLKAGNKVTSSESSLLTSELRSKREQYKQ	240	
Db	181	STGLVSNVKPLSLVNOAKKETYMKLVPSPDVLKAGNKVTSSESSLLTSELRSKREQYKQ	240	
Qy	241	AFPSQLKKQESSKSLKKVIAALSNPKATSSSPAHPKQTLNHNHPNPLTNALLGNHPNG	300	
Db	241	AFPSQLKKQESSKSLKKVIAALSNPKATSSSPAHPKQTLNHNHPNPLTNALLGNHPNG	300	
Qy	301	VIQSVTOEAPLATTTKTMOSKINENIAAASSTPSSPVNLSVSGRTPCNGOTPVMPSPAS	360	
Db	301	VIQSVTOEAPLATTTKTMOSKINENIAAASSTPSSPVNLSVSGRTPCNGOTPVMPSPAS	360	
Qy	361	PILHSQGKEKAVANNVNPVKTQHHSHPAKSLVQFRGTDSDIPSSKDSNSDEDEEDDE	420	
Db	361	PILHSQGKEKAVANNVNPVKTQHHSHPAKSLVQFRGTDSDIPSSKDSNSDEDEEDDE	420	
Qy	421	EEDEDEDEDESDSQSESDSNSSEDTGSEEEEDDDDKQODESDSDTEGEKTSMLNKTT	480	
Db	421	EEDEDEDEDESDSQSESDSNSSEDTGSEEEEDDDDKQODESDSDTEGEKTSMLNKTT	480	
Qy	481	SSVKSPMSLTHGSTPRNLHIAKPGSAPAAALCSQSPAFGLGTSSTLTSSPHSGTSSKR	540	
Db	481	SSVKSPMSLTHGSTPRNLHIAKPGSAPAAALCSQSPAFGLGTSSTLTSSPHSGTSSKR	540	
Qy	541	RRVTDRELRIPLYGHWORETRIRNFGRLQGEVAYVYAPCGKKLROYPEVIKYSRNGIM	600	
Db	541	RRVTDRELRIPLYGHWORETRIRNFGRLQGEVAYVYAPCGKKLROYPEVIKYSRNGIM	600	

QY 601 DISRDNFSFSAKIRVGDYFVARDGPOEMQWCLLKEEDVI PRIRAMEGRGRPPNDPDRORA 660
DB 601 DISRDNFSFSAKIRVGDYFVARDGPOEMQWCLLKEEDVI PRIRAMEGRGRPPNDPDRORA 660
QY 661 REESRMRRKGRPPNVGNAEFLDNADAKLLRLKLAQAEIARQAQIKLLRLKQKQEQARVA 720
DB 661 REESRMRRKGRPPNVGNAEFLDNADAKLLRLKLAQAEIARQAQIKLLRLKQKQEQARVA 720
QY 721 KEAKKQQAIIAAEERKQKEQIKMKQKEKIKRIQOIRMEKELRAQOILEAKKKKKEBAA 780
DB 721 KEAKKQQAIIAAEERKQKEQIKMKQKEKIKRIQOIRMEKELRAQOILEAKKKKKEBAA 780
QY 781 NAKLLEAEKRIKEEMRRCQAVLLKHQERERRRHHMLKWEARKKAEEKERLQEKRD 840
DB 781 NAKLLEAEKRIKEEMRRCQAVLLKHQERERRRHHMLKWEARKKAEEKERLQEKRD 840
QY 841 EKRLNKKERLQORRIELEMAKELKPNEDMCLADQKPLPELPRIPGLVLSGSTFSDCLMV 900
DB 841 EKRLNKKERLQORRIELEMAKELKPNEDMCLADQKPLPELPRIPGLVLSGSTFSDCLMV 900
QY 901 VOFLRNFQKVLGFVDNIIDVNLVLOEGLLNIIGDSMGVQDILLVRLLSAAVCDPGLITGY 960
DB 901 VOFLRNFQKVLGFVDNIIDVNLVLOEGLLNIIGDSMGVQDILLVRLLSAAVCDPGLITGY 960
QY 961 KAKTALGHEHLLNVGNRRDNVSEIILQIFMEAHCQGTTELTSKTKAFOAHTTAPAKASVLAF 1020
DB 961 KAKTALGHEHLLNVGNRRDNVSEIILQIFMEAHCQGTTELTSKTKAFOAHTTAPAKASVLAF 1020
QY 1021 LINELACSKSVVSEIDKNIDYMSNLRDKVVEGKRLKRIIHAKTGKTGRTSGGIDIGE 1080
DB 1021 LINELACSKSVVSEIDKNIDYMSNLRDKVVEGKRLKRIIHAKTGKTGRTSGGIDIGE 1080
QY 1081 EOHPLGTPTPKRRRRKGGSDYDDDDDDSDQDDEDEDEEKEDQKHKTTDICEDED 1140
DB 1081 EOHPLGTPTPKRRRRKGGSDYDDDDDDSDQDDEDEDEEKEDQKHKTTDICEDED 1140
QY 1141 EGDQAASVEELEKQIEKLSKQOSQYRRLFPASHLSRSMVEGPDYRRYRWILPRCCGIF 1200
DB 1141 EGDQAASVEELEKQIEKLSKQOSQYRRLFPASHLSRSMVEGPDYRRYRWILPRCCGIF 1200
QY 1201 VEGMESGGLBEIAKEREKLLKAESVQIKEEMFETSGDSLNCSTNDHCEQKEDLKEKONT 1260
DB 1201 VEGMESGGLBEIAKEREKLLKAESVQIKEEMFETSGDSLNCSTNDHCEQKEDLKEKONT 1260
QY 1261 NLFLOKPGSFKLSKLLVAVAKMPESEVWTPKPNAGANGCTLSYQNSGKHSGLSVQSTAT 1320
DB 1261 NLFLOKPGSFKLSKLLVAVAKMPESEVWTPKPNAGANGCTLSYQNSGKHSGLSVQSTAT 1320
QY 1321 QSNVEKADSNLNFNTSGSGPGKFYSPLPNDQLLTKLTEKNQWFSLLPRTPCDDTSLTHA 1380
DB 1321 QSNVEKADSNLNFNTSGSGPGKFYSPLPNDQLLTKLTEKNQWFSLLPRTPCDDTSLTHA 1380
QY 1381 DMSTASLVTPOSPPSKSPSTPAPLGSSAQNVPVGLNPFALSPLOKVGGSVMGLOFCGW 1440
DB 1381 DMSTASLVTPOSPPSKSPSTPAPLGSSAQNVPVGLNPFALSPLOKVGGSVMGLOFCGW 1440
QY 1441 PTGVVTSNIPPTLSVPSLGLGLGSEGNGNSFLTSNVASSKSESFPVONEKATSAQPAAV 1500
DB 1441 PTGVVTSNIPPTLSVPSLGLGLGSEGNGNSFLTSNVASSKSESFPVONEKATSAQPAAV 1500
QY 1501 EVAKPVDPPSPKPIPEEMQFCGWRIIDPEDLKALLKVLHLRGIREKALOKQIQKHLDYIT 1560
DB 1501 EVAKPVDPPSPKPIPEEMQFCGWRIIDPEDLKALLKVLHLRGIREKALOKQIQKHLDYIT 1560
QY 1561 QACLKKNKDVAIIELNENEENQVTRDIVENWVSVEEQAMEMDLVLSQVQVEDLERRVASASLQ 1620
DB 1561 QACLKKNKDVAIIELNENEENQVTRDIVENWVSVEEQAMEMDLVLSQVQVEDLERRVASASLQ 1620
QY 1621 VKGMWCPBPASEREDLVYFEHKSFTKLCCKHDGFTGDESSAHALERKSNPLDIAVTR 1680
DB 1621 VKGMWCPBPASEREDLVYFEHKSFTKLCCKHDGFTGDESSAHALERKSNPLDIAVTR 1680

QY 1681 LADLERNIERRIEDIIAPGLRVWRRLSEARSAQAVALCIIQLOKSIWAKSIMKVYCOI 1740
DB 1681 LADLERNIERRIEDIIAPGLRVWRRLSEARSAQAVALCIIQLOKSIWAKSIMKVYCOI 1740
QY 1741 CRKGDNEELLLLCDGCDKGCHTYCHRPKITTIPDGDWFCPACIAKASGQTLIKKLHVKG 1800
DB 1741 CRKGDNEELLLLCDGCDKGCHTYCHRPKITTIPDGDWFCPACIAKASGQTLIKKLHVKG 1800
QY 1801 KKTNESKGGKVVTLTGDTEDDSASTSSSLKRGKNDLOKRMDENTSINLSKQESFTSVK 1860
DB 1801 KKTNESKGGKVVTLTGDTEDDSASTSSSLKRGKNDLOKRMDENTSINLSKQESFTSVK 1860
QY 1861 KPRKDDSKDALCSMILTEMETHEDAWPFLLPVNLKLVPGYKVIKKPMDPSTIREKLSS 1920
DB 1861 KPRKDDSKDALCSMILTEMETHEDAWPFLLPVNLKLVPGYKVIKKPMDPSTIREKLSS 1920
QY 1921 GOYPNLETFTALDVLRLVFDNCETFNEDDSDIGRAGHNNMRKYFEKKWTDTFKVS 1972
DB 1921 GOYPNLETFTALDVLRLVFDNCETFNEDDSDIGRAGHNNMRKYFEKKWTDTFKVS 1972
RESULT 3
AAW81171
ID AAW81171 standard; protein; 1972 AA.
XX AC AAW81171;
XX AC AAW81171;
DT 05-MAY-1999 (first entry)
XX Human BAZ2-beta protein.
DE Human BAZ2-beta protein.
KW Transcriptional regulator; transcription; BAZ1-alpha; bromodomain; BAZ2;
KW atypical zinc finger; testis; human; tumour; BAZ1-beta; BAZ2-alpha; drug;
KW BAZ2-beta; treatment; cancer; proliferative disorder; screening.
XX Homo sapiens.
XX Homo sapiens.
PN W09847920-A1.
XX W09847920-A1.
PD 29-OCT-1998.
XX 29-OCT-1998.
PF 17-APR-1998; 98WO-JP001783.
XX 17-APR-1998; 98WO-JP001783.
PR 18-APR-1997; 97JP-00116570.
XX 18-APR-1997; 97JP-00116570.
PR 24-OCT-1997; 97JP-00310027.
XX 24-OCT-1997; 97JP-00310027.
PA (CHUS) CHUGAI RES INST MOLECULAR MEDICINE INC.
XX (CHUS) CHUGAI RES INST MOLECULAR MEDICINE INC.
PI Jones MH;
XX Jones MH;
XX WPI; 1998-583603/49.
DR N-PSDB; AAV68403.
XX N-PSDB; AAV68403.
PT Transcriptional regulator gene family containing bromodomain - may be
PT expressed in testis tissue and is useful for treatment of cancer and
PT other proliferative disorders.
XX Claim 1; Page 100-116; 187pp; Japanese.
PS Claim 1; Page 100-116; 187pp; Japanese.
XX This sequence represents the human BAZ2-beta protein, a member of a
CC family of transcriptional regulator genes containing a bromodomain (BAZ,
CC Bromodomain with Atypical Zinc finger) which are expressed specifically
CC in testis tissue and also in certain tumour lines. Transgenic cells may
CC be used for the preparation of the BAZ1-alpha, BAZ1-beta, BAZ2-alpha and
CC BAZ2-beta proteins. These proteins can be used in the treatment of cancer
CC and other proliferative disorders, and in screening of compounds for
CC their binding ability to the expression products (e.g. for use as drugs
CC by modulation of transcriptional regulation)
XX Sequence 1972 AA;
SQ Sequence 1972 AA;

Query Match 99.9%; Score 10131; DB 2; Length 1972;
Best Local Similarity 99.9%; Pred. No. 0;

Matches 1971; Conservative 0; Mismatches 1; Indels 0; Gaps 0;			
Qy	1	MGQTSTSGGNRKNQSQKNQPLDARVDKI	KDKKPRKXWESSNSDSGTSSTDS 60
Db	1	MGQTSTSGGNRKNQSQKNQPLDARVDKI	KDKKPRKXWESSNSDSGTSSTDS 60
Qy	61	SEGISSDDLEDEEEDQSI	TESEDDSDSESAQHSNNVLLHGISDPKADGQKA 120
Db	61	SEGISSDDLEDEEEDQSI	TESEDDSDSESAQHSNNVLLHGISDPKADGQKA 120
Qy	121	TEKAQEKRIHQPLPLAFESQTHS	FQSQOQPOVLSQQLPFIQSSQAKESVNGKTSVIQ 180
Db	121	TEKAQEKRIHQPLPLAFESQTHS	FQSQOQPOVLSQQLPFIQSSQAKESVNGKTSVIQ 180
Qy	181	STGLVSNVPLSLVNOAKKETYMKLIV	PSDPVLKAGNKNTSESSLLTSELRSKREYKQ 240
Db	181	STGLVSNVPLSLVNOAKKETYMKLIV	PSDPVLKAGNKNTSESSLLTSELRSKREYKQ 240
Qy	241	APPSOLKKQESSKSLKVVIAALSNPKATSS	PAHPKQTLNHNHPPLTNALLGNHQPN 300
Db	241	APPSOLKKQESSKSLKVVIAALSNPKATSS	PAHPKQTLNHNHPPLTNALLGNHQPN 300
Qy	301	VTSQVITQEAPLATTKTKMQSKINENI	AAASGTFPSSPVNLSTSGRRTPCNOTPVMPAS 360
Db	301	VTSQVITQEAPLATTKTKMQSKINENI	AAASGTFPSSPVNLSTSGRRTPCNOTPVMPAS 360
Qy	361	PILHSQKGEKAVSNVNPVKTOHSHPAKSL	VEQPRGTDSDIPSSKOSDSNEDEEDE 420
Db	361	PILHSQKGEKAVSNVNPVKTOHSHPAKSL	VEQPRGTDSDIPSSKOSDSNEDEEDE 420
Qy	421	EDEDEDEDESDSQSESDSSESDTEGS	EDEDDDDKQDSDSDTEGEKTSMLKNTT 480
Db	421	EDEDEDEDESDSQSESDSSESDTEGS	EDEDDDDKQDSDSDTEGEKTSMLKNTT 480
Qy	481	SVKQPSNLTGHSPTRNLIHAKPGSAPAAL	CSQSQSPAFLTGTSSTLTSPPHSGTSKR 540
Db	481	SVKQPSNLTGHSPTRNLIHAKPGSAPAAL	CSQSQSPAFLTGTSSTLTSPPHSGTSKR 540
Qy	541	RRVTDERELRIIPLEYGWQRETRIRNFG	RLOGEVAYAPCGKKLRQYPEVVKYLSRNGIM 600
Db	541	RRVTDERELRIIPLEYGWQRETRIRNFG	RLOGEVAYAPCGKKLRQYPEVVKYLSRNGIM 600
Qy	601	DISRDNFSAKIRVGDFFYEARDGQOEMQ	WCLLKEEDVIPRIAMEGRGRPPNPDRORA 660
Db	601	DISRDNFSAKIRVGDFFYEARDGQOEMQ	WCLLKEEDVIPRIAMEGRGRPPNPDRORA 660
Qy	661	REESRMRKGRPPNVGNAEFLDNADAKLL	RKLAQETARAAQIKLLRKLOQOQARVA 720
Db	661	REESRMRKGRPPNVGNAEFLDNADAKLL	RKLAQETARAAQIKLLRKLOQOQARVA 720
Qy	721	KEAKQQAIAAEERKQKEQIKIMKQEQEKI	KRIQIIRMEKELRAQQILAEAKKKKEBA 780
Db	721	KEAKQQAIAAEERKQKEQIKIMKQEQEKI	KRIQIIRMEKELRAQQILAEAKKKKEBA 780
Qy	781	NAKLLAEAKRIKEKEMRROQAALLKHQER	RERRRQHMMLKMAEARKKAEERLKOEKRD 840
Db	781	NAKLLAEAKRIKEKEMRROQAALLKHQER	RERRRQHMMLKMAEARKKAEERLKOEKRD 840
Qy	841	EKRLNKRLEQRRLLENAKELKKPNEDM	CLADOKPLPELPRIPGLVLSGTSFSDCLMV 900
Db	841	EKRLNKRLEQRRLLENAKELKKPNEDM	CLADOKPLPELPRIPGLVLSGTSFSDCLMV 900
Qy	901	VQFLRNFGLVGFVDNIDVPNLSVLQEG	LLNIGDSMGEVQDILLVLLSAAVCDPLITGY 960
Db	901	VQFLRNFGLVGFVDNIDVPNLSVLQEG	LLNIGDSMGEVQDILLVLLSAAVCDPLITGY 960
Qy	961	KAKTALGEHLNVGNRNVNSBILQIFME	AHCGQTELTSKTKAPQAHTPAQKASVLAF 1020
Db	961	KAKTALGEHLNVGNRNVNSBILQIFME	AHCGQTELTSKTKAPQAHTPAQKASVLAF 1020
Qy	1021	LINELACSKSVSEIDKNDIDYNSNRRDK	WVVEGKLRLRIIHAKTGKRTSGGIDLGE 1080
Db	1021	LINELACSKSVSEIDKNDIDYNSNRRDK	WVVEGKLRLRIIHAKTGKRTSGGIDLGE 1080

Qy	1081	EQHPIGTPTPGRRKRRKGGSDSYDDDD	DDSDSDQGGDEDEDEKEDQKKTDCI	CEDED 1140
Db	1081	EQHPIGTPTPGRRKRRKGGSDSYDDDD	DDSDSDQGGDEDEDEKEDQKKTDCI	CEDED 1140
Qy	1141	EGDQAASVEELEKQTEKLSKOOSQYRR	KLFDASHLSRVMFPGPDRVRRYWLPR	CGGIP 1200
Db	1141	EGDQAASVEELEKQTEKLSKOOSQYRR	KLFDASHLSRVMFPGPDRVRRYWLPR	CGGIP 1200
Qy	1201	VEGMESGEGLEBIAKEREKLLKKAESVQ	IKKEEMFETSGDSLNCNTDHCQKEDL	KEKONT 1260
Db	1201	VEGMESGEGLEBIAKEREKLLKKAESVQ	IKKEEMFETSGDSLNCNTDHCQKEDL	KEKONT 1260
Qy	1261	NLFLOKPGSFSKLSLLEVAKMPPESEV	MTPKPNAGANGCTLSYQNSGHSLGS	VQSTAT 1320
Db	1261	NLFLOKPGSFSKLSLLEVAKMPPESEV	MTPKPNAGANGCTLSYQNSGHSLGS	VQSTAT 1320
Qy	1321	QSNVEKASNNLFTNGSSGPGKFYSP	LPNDQLLTKLTTEKNRWFSLPRTT	PCDDTSLTHA 1380
Db	1321	QSNVEKASNNLFTNGSSGPGKFYSP	LPNDQLLTKLTTEKNRWFSLPRTT	PCDDTSLTHA 1380
Qy	1381	DMSTASLVTPOSQPPSKSPSTPAPLGS	SAQNPVGLNPFALSPLOVKGVSMM	QFCGW 1440
Db	1381	DMSTASLVTPOSQPPSKSPSTPAPLGS	SAQNPVGLNPFALSPLOVKGVSMM	QFCGW 1440
Qy	1441	PTGVVTSNIPFTLSVPSLGSGLGSEGN	SFLTNSVASSSESVPQNEKATSQA	PAAV 1500
Db	1441	PTGVVTSNIPFTLSVPSLGSGLGSEGN	SFLTNSVASSSESVPQNEKATSQA	PAAV 1500
Qy	1501	EYAKVDPDPSPKPIPEEMQFGWRIID	DPEDLKALLKVLHLRGIREKALQK	IOKHLDYIT 1560
Db	1501	EYAKVDPDPSPKPIPEEMQFGWRIID	DPEDLKALLKVLHLRGIREKALQK	IOKHLDYIT 1560
Qy	1561	QACLKXKQVAIIELNENEENOVTRDI	VENWSVEQAMEMDLSVLOQVEDL	ERRVASASIQ 1620
Db	1561	QACLKXKQVAIIELNENEENOVTRDI	VENWSVEQAMEMDLSVLOQVEDL	ERRVASASIQ 1620
Qy	1621	VKGWMCPEPASERDLVYFEHKSFTKL	CKEHGDEFTGEDESSAHLERKSD	NPLDIAVTR 1680
Db	1621	VKGWMCPEPASERDLVYFEHKSFTKL	CKEHGDEFTGEDESSAHLERKSD	NPLDIAVTR 1680
Qy	1681	LADLERNTERRIEEDIAFLRWRRAL	SEARSAAQVALCIIQOLQKSI	AWKSIWKYCOI 1740
Db	1681	LADLERNTERRIEEDIAFLRWRRAL	SEARSAAQVALCIIQOLQKSI	AWKSIWKYCOI 1740
Qy	1741	CRKGNBELLALLCDGCDKGCHTYCH	RPKITITIPDGFWFCPACIAKASG	QTLKIKLHVKG 1800
Db	1741	CRKGNBELLALLCDGCDKGCHTYCH	RPKITITIPDGFWFCPACIAKASG	QTLKIKLHVKG 1800
Qy	1801	KKTNESKKGKVTTLTGDTDEDESD	SSTSSSKRGNKDQKRMEENTS	INLSKQESFTSYK 1860
Db	1801	KKTNESKKGKVTTLTGDTDEDESD	SSTSSSKRGNKDQKRMEENTS	INLSKQESFTSYK 1860
Qy	1861	KPKRDDSXLALCSMILTEMETHED	AWPFLIPVNLKLVPGYKVKIKK	PMDFSTIREKLSS 1920
Db	1861	KPKRDDSXLALCSMILTEMETHED	AWPFLIPVNLKLVPGYKVKIKK	PMDFSTIREKLSS 1920
Qy	1921	GOYPNLETALDVLVFDNCETFNED	SDSDIGRAGHNMRYPEKKTDT	FKVS 1972
Db	1921	GOYPNLETALDVLVFDNCETFNED	SDSDIGRAGHNMRYPEKKTDT	FKVS 1972

RESULT 4
AAM93402
ID AAM93402 standard; protein; 796 AA.
XX
AC AAM93402;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human polypeptide, SEQ ID NO: 3004.
XX
KW Human; full length cDNA; cDNA synthesis; oligo-capping.

CC glycoproteins and transcription. The present invention describes a method
CC for efficiently cloning a full length human cDNA from both the 5' and 3',
CC ends using the oligo-capping method. This polypeptide sequence is a full
CC length human protein of the invention.

XX SQ Sequence 796 AA;
Query Match 41.0%; Score 4153; DB 8; Length 796;
Best Local Similarity 99.5%; Pred. No. 8.5e-265;
Matches 792; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
QY 1042 MSNLRDQWVVEGKRLRIHAKTKGRDTSGGIDGEEQHPGLTPTGRRKRGKGS 1101
DB 1 MSNLRDQWVVEGKRLRIHAKTKGRDTSGGIDGEEQHPGLTPTGRRKRGKGS 60
QY 1102 DYDDDDDDSDQGGDEDEDEDKQKTKTDICEDEDEGQAAASVEELEKQIEKSKQ 1161
DB 61 DYDDDDDDSDQGGDEDEDEDKQKTKTDICEDEDEGQAAASVEELEKQIEKSKQ 120
QY 1162 QSQYRKLFDAHSLSRSMFGRYRRYWLPRCGGIFVEGMEGEGLEIAKEREK 1221
DB 121 QSQYRKLFDAHSLSRSMFGRYRRYWLPRCGGIFVEGMEGEGLEIAKEREK 180
QY 1222 KAESVQIKKEEMFETSGDSLNCSTNTHCEQEDLKEKNTNLPLOKPGSFSLKLELVAK 1281
DB 181 KAESVQIKKEEMFETSGDSLNCSTNTHCEQEDLKEKNTNLPLOKPGSFSLKLELVAK 240
QY 1282 MPPESEVMTPKPNAGANGCTLSYQNSGKHSLSGVSQSTQSNVEKADNNLNTGSSGPG 1341
DB 241 MPPESEVMTPKPNAGANGCTLSYQNSGKHSLSGVSQSTQSNVEKADNNLNTGSSGPG 300
QY 1342 KFYSLPNDQLLKTLEKNRQWFSLLPRTPCDDTSLTHADMTASLVTPOQPPSKSPSP 1401
DB 301 KFYSLPNDQLLKTLEKNRQWFSLLPRTPCDDTSLTHADMTASLVTPOQPPSKSPSP 360
QY 1402 TPAPLGSSAQNVPGLNPFALSPLOKGVSMGLQFCGWPTGVTSNPTLSPVSLGSG 1461
DB 361 TPAPLGSSAQNVPGLNPFALSPLOKGVSMGLQFCGWPTGVTSNPTLSPVSLGSG 420
QY 1462 LGLSGNGSFLTSNVASKSPSPVQNEKATSAQAAVEAKPVDFPSPKPIPEMOFG 1521
DB 421 LGLSGNGSFLTSNVASKSPSPVQNEKATSAQAAVEAKPVDFPSPKPIPEMOFG 480
QY 1522 WRRITDPEDLKALLKVLHLRGIREKALQKIQKHLDIYITQACLKNKQVAIIELNENE 1581
DB 481 WRRITDPEDLKALLKVLHLRGIREKALQKIQKHLDIYITQACLKNKQVAIIELNENE 540
QY 1582 VTRDIVENWSVEEQAMENDLSVLQVEDLERRVASASLQVKGWMCPEPASEREDLVYFEH 1641
DB 541 VTRDIVENWSVEEQAMENDLSVLQVEDLERRVASASLQVKGWMCPEPASEREDLVYFEH 600
QY 1642 KSFTKLCHEHDEFTGEDESSAHALERKSDNPLDIAVTRLADLERNIERRIEEDIAPGLR 1701
DB 601 KSFTKLCHEHDEFTGEDESSAHALERKSDNPLDIAVTRLADLERNIERRIEEDIAPGLR 660
QY 1702 VRRALSEARSAAQVALCIQLOKSIAMEKSIIMKYICQICRKGDNELLLCDGCKGCH 1761
DB 661 VRRALSEARSAAQVALCIQLOKSIAMEKSIIMKYICQICRKGDNELLLCDGCKGCH 720
QY 1762 TYCHRPKTTTTPDGWFPCPACIAKASGQTLKIKLHVKGKTNESKKGKVTLTGDTDE 1821
DB 721 TYCHRPKTTTTPDGWFPCPACIAKASGQTLKIKLHVKGKTNESKKGKVTLTGDTDE 780
QY 1822 DSASTSSSLKRGKNDL 1837
DB 781 DSASTSSSLKRGKNDL 796
RESULT 6
AAB93073 standard; protein; 708 AA.
XX AC AAB93073;

XX 26-JUN-2001 (first entry)
DT Human protein sequence SEQ ID NO:11898.
DE Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX Homo sapiens.
KW EP1074617-A2.
XX 07-FEB-2001.
XX 28-JUL-2000; 2000EP-00116126.
XX 29-JUL-1999; 99JP-00248036.
PR 27-AUG-1999; 99JP-00300253.
PR 11-JAN-2000; 2000JP-00118776.
PR 02-MAY-2000; 2000JP-00183767.
PR 09-JUN-2000; 2000JP-00241899.
XX (HELI-) HELIX RES INST.
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI; 2001-318749/34.
XX Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.
XX Claim 8; SEQ ID NO 11898; 2537pp + Sequence Listing; English.
XX The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention
XX Sequence 708 AA;
Query Match 35.1%; Score 3559; DB 4; Length 708;
Best Local Similarity 94.9%; Pred. No. 1.1e-225;
Matches 704; Conservative 2; Mismatches 2; Indels 34; Gaps 1;
QY 474 MLNKTTSVKSPSMSLTGHSHTPRNLHIAPGAPAPALCSSESQSPALGTSSTLTSSP 533
DB 1 MLNKTTSVKSPSMSLTGHSHTPRNLHIAPGAPAPALCSSESQSPALGTSSTLTSSP 60
QY 534 HSGTSKRRRVTDRELRIPLRYGQRETRINFGRLQGEVAYVAPCGKRLQYPEVTKY 593
DB 61 HSGTSKRRRVTDRELRIPLRYGQRETRINFGRLQGEVAYVAPCGKRLQYPEVTKY 119
QY 594 LSRNGIMDISRDNFSFSAKIRVGDGFYEARDGQENQWCLKEEDVIPRAMEGRGRPP 653

```
Db 120 -----GQWCLLKEDVPIPRAMEGRGRPP 146
Qy 654 NPDORAREESMRRRKGRPPNVGNAEFLDNADAKLRLKLAQETARQAAQIKLRLKLOK 713
Db 147 NPDORAREESMRRRKGRPPNVGNAEFLDNADAKLRLKLAQETARQAAQIKLRLKLOK 206
Qy 714 QEQARVAEAKQQAIAEAEKKEKQEI KIMKQOEKIKRTOOIRMEKELRAQOILEAKK 773
Db 207 QEQARVAEAKQQAIAEAEKKEKQEI KIMKQOEKIKRTOOIRMEKELRAQOILEAKK 266
Qy 774 KKKEEAAAKLLEAEKRIKEKEMRQQAVALLKHOERERRRHMMMLKMAEARKKAEKEER 833
Db 267 KKKEEAAAKLLEAEKRIKEKEMRQQAVALLKHOERERRRHMMMLKMAEARKKAEKEER 326
Qy 834 LKQEKREDEKRLNKKERKLEQRLELEMAKELKKPNEDMCLADQKPLPELPRIPGLVLSGST 893
Db 327 LKQEKREDEKRLNKKERKLEQRLELEMAKELKKPNEDMCLADQKPLPELPRIPGLVLSGST 386
Qy 894 FSDCLMVVQFLRNFGKVLGFDVNDVNLVLVLOEGLNIGDSMGHVODLLVRLLSAAVCD 953
Db 387 FSDCLMVVQFLRNFGKVLGFDVNDVNLVLVLOEGLNIGDSMGHVODLLVRLLSAAVCD 446
Qy 954 PGLITGYKAKTALGHEHLNVGNRDNVSEILQIFMEAHCGQOTELTESLTKAFQAHTPAQ 1013
Db 447 PGLITGYKAKTALGHEHLNVGNRDNVSEILQIFMEAHCGQOTELTESLTKAFQAHTPAQ 506
Qy 1014 KASVLAFILNELACSKSVSEIDKNIDYMSNLRDKWVVEGKRLKRIIHAKTGKRDTS 1073
Db 507 KASVLAFILNELACSKSVSEIDKNIDYMSNLRDKWVVEGKRLKRIIHAKTGKRDTS 566
Qy 1074 GGIDLGEHQHPLGTPTGRKRRKGGSDYDDDDDDSDGDEDEDEDEKDGKGGKT 1133
Db 567 GGIDLGEHQHPLGTPTGRKRRKGGSDYDDDDDDSDGDEDEDEDEKDGKGGKT 626
Qy 1134 DICEDEDEGQQAASVEELEKQIEKLSKQSOVRRKLFDAHSLSRVMPGDRYRRRWIL 1193
Db 627 DICEDEDEGQQAASVEELEKQIEKLSKQSOVRRKLFDAHSLSRVMPGDRYRRRWIL 686
Qy 1194 PRCGGIFVEGMESGEGLEEIAK 1215
Db 687 PRCGGIFVEGMESGEGLEEIAK 708

RESULT 7
ID ADR09338
AC ADR09338 standard; protein; 589 AA.
XX ADR09338;
DT 04-NOV-2004 (first entry)
DE Human protein useful for treating neurological disease Seq 2844.
KW human; oligo-capping method; diagnostic marker; gene therapy;
KW osteoporosis; neurological disease; Alzheimer's disease;
KW Parkinson's disease; dementia; short memory; cancer;
KW sense or motor function; emotional reaction; fear response; panic;
KW osteopathic; neuroprotective; nootropic; antiparkinsonian; cyostatic;
KW tranquiliser.
OS Homo sapiens.
FN EP1447413-A2.
XX 18-AUG-2004.
XX 12-FEB-2004; 2004EP-00003145.
XX 14-FEB-2003; 2003JP-00102207.
PR 09-MAY-2003; 2003JP-00131452.
XX (REAS-) RES ASSOC BIOTECHNOLOGY.
```

```
XX Isogai T, Yamamoto J, Nishikawa T, Isono Y, Sugiyama T, Otsuki T;
PI Wakamatsu A, Ishii S, Nagai K, Irie R;
XX WPI; 2004-583265/57.
DR N-PSDB; ADR07382.
XX
PT New 1995 cDNA, useful for treating osteoporosis, neurological diseases,
PT Alzheimer's diseases, Parkinson's diseases, dementia and various cancers.
XX
PS Claim 1; SEQ ID NO 2844; 2686pp; English.
XX
XX This invention relates to novel, isolated full length human cDNA
XX molecules and the encoded proteins thereof. Specifically, it refers to
XX cDNA clones obtained by an oligo-capping method, where none of these
XX clones are identical to any known human mRNAs. The present invention
XX describes an immunoassay to identify agonists and antagonists, as well as
XX antibodies, antisense molecules and siRNAs that can all be used to bind
XX to and modulate expression of the cDNA molecules. As such, these
XX molecules are useful for diagnostic markers or therapeutic targets for
XX the various diseases or morbid states. In particular, they are useful in
XX gene therapy for treating osteoporosis, neurological disease, Alzheimer's
XX disease, Parkinson's disease, dementia, short memory and various cancers,
XX as well as for maintaining equilibrium of sense or motor function, and
XX for treating emotional reaction, fear response and panic. Accordingly,
XX they exhibit osteopathic, neuroprotective, nootropic, antiparkinsonian,
XX cyostatic and tranquiliser activities. This polypeptide is a protein
XX encoded by a full length human cDNA sequence of the invention. NOTE: This
XX sequence is not given in the sequence listing of the specification but
XX can be obtained on CD-ROM from the European Patent Office, Vienna Sub-
XX office.
XX
XX Sequence 589 AA;
SQ
Query Match 28.6%; Score 2903; DB 8; Length 589;
Best Local Similarity 99.5%; Pred. No. 1.7e-182;
Matches 573; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
Qy 1 MGQTKSTSSGGNRKNCQEQSKNQPLDARVDKIKDKPKRKAMSSNSDSGTSSTDS 60
Db 1 MGQTKSTSSGGNRKNCQEQSKNQPLDARVDKIKDKPKRKAMSSNSDSGTSSTDS 60
Qy 61 SEGISSSDSDLEDEDEEDQSI ESEDDSDSEAEQHSNNQVLLHG1SDPKADGQKA 120
Db 61 SEGISSSDSDLEDEDEEDQSI ESEDDSDSEAEQHSNNQVLLHG1SDPKADGQKA 120
Qy 121 TEKAQEKRIHQPLAPFESQTHSFQSQKQPVLSQQLPFIFQSSQAKEESVNKHTSVIQ 180
Db 121 TEKAQEKRIHQPLAPFESQTHSFQSQKQPVLSQQLPFIFQSSQAKEESVNKHTSVIQ 180
Qy 181 STGLVSNVKNPLSLVNQAKKETMYMKLIVPSPDVLKAGNKVTSEESSLLTSELRSKREYKQ 240
Db 181 STGLVSNVKNPLSLVNQAKKETMYMKLIVPSPDVLKAGNKVTSEESSLLTSELRSKREYKQ 240
Qy 241 APPSOLKQESSKSLKKVIAALSNPKATSSSPAHPKOTLENNHPNPFLLTALLGNHQPNG 300
Db 241 APPSOLKQESSKSLKKVIAALSNPKATSSSPAHPKOTLENNHPNPFLLTALLGNHQPNG 300
Qy 301 VIQSVIQAEPALATTTKTMQSKINENIAAASSTPFSSPVLNSTSGRRTPGNQTPVMPSPAS 360
Db 301 VIQSVIQAEPALATTTKTMQSKINENIAAASSTPFSSPVLNSTSGRRTPGNQTPVMPSPAS 360
Qy 361 PILHSQKEKAVSNVNPVKTQHSHHPAKSLIVEQFRGTDSDIPSSKSEDSNEDEEEDDE 420
Db 361 PILHSQKEKAVSNVNPVKTQHSHHPAKSLIVEQFRGTDSDIPSSKSEDSNEDEEEDDE 420
Qy 421 EDEDEDEDESDSDSQSESDNSSESDTEGSEERDDDDKDDSDSDTEGKTSMLNKTT 480
Db 421 EDEDEDEDESDSDSQSESDNSSESDTEGSEERDDDDKDDSDSDTEGKTSMLNKTT 480
Qy 481 SSVKSPSM1TGHSTPRNLHIKAPGSAPALACSESQSPAF1GTSSTLTSSPHSGTSKR 540
Db 481 SSVKSPSM1TGHSTPRNLHIKAPGSAPALACSESQSPAF1GTSSTLTSSPHSGTSKR 540
```


QY 261 ALSNPKATSSPAHPKQTLNHNHPNPLTNALLGNHQPNGVITQSVTQEAFLALTTKTKMQ 320
Db 238 VLV-PDPTVSCUDDP-----SHLPD-----QLEDTP----- 263
QY 321 SKINENIAAASPPFSSPNVNSTSGRRTPGNQTPVMPSPASPILHSGCKEKAVNNVNVK 380
Db 264 -----LSEDSLFPNSLAPEPVSGGLYCIDOTELM-----GAEKLPLEDSPI 307
QY 381 T-----QHSHHPAKSLVVEFRGTDSIPSSKDSOSNEDEDEDEDEDEDEDESDS 435
Db 308 SALDCPSLNNATAFSLADDSTSTSFASPTSPVILGESVLQDNSFDLNNNGSDAEQEM 367
QY 436 QSESNSNSDTEGSEEDDDDDKQDSDSDTEGEKTSMLNKTTS----- 481
Db 368 ETQSSDFFPSLTOPA-----PDQSTIQLHPATSPAVSPTSPAVSLVVSAAASPE 418
QY 482 -----SVKSPSMS-----LTGHSPTPRNLHIKAPGSA-PAA 511
Db 419 ISPEVCPAASTVVSFAVSVSPASSAVLPVAVSLEVLPTASTVSPKASPTSPAAAFPTA 478
QY 512 LCSESQSPALGTSST-----LTSSPHSGTSRRRRTVDRELRIPLEYGQRETRIR 564
Db 479 SPANKDVSSFLETTADVEITCEGLTAS-GSGDVMERRIATPEEVRLPLQHGWRREVRIK 537
QY 565 NFGGRLOGEVAYAPCGKKLRQYPEVIKYLRSNGIMDISRDNPSPSAKIRVGDFEARDG 624
Db 538 KGSHRWQGTWYGPCGKMKQFPEVIKYLRSNVVHVSRRHFSFSPRMPVGDFFPEERT 597
QY 625 POEMOWCLLKEDVTPRIAMGRGRPPNPDRORARESRMRGRPPNVGNAEFLDN 684
Db 598 PGLQWQLSABEISRIQITGKGRPRNTKATKEVPKVRGRGRPPKVIITELNKK 657
QY 685 ADALKRLKQAEIARQAQAIKLLRKLQKQEARVAKEAKQ-----QAIMAAEBK 735
Db 658 TDRNPLKLEAQE-----TLNEEDKAKIAKSKKMKVQKQVQGECCOTTIQGAR 705
QY 736 RKQEQIKIMKQEKIKRIQITRMEKELRAQOILEAKKKKBEANAKLLEAKRIKKE 795
Db 706 NKROBTSLKQK-----EAKKSKKEGKTK-----QELKKEK- 739
QY 796 MRRQAVLLKHQERRRRQHMLKAMEARKAEKELKQEKDEKRLNKEKLEORRL 855
Db 740 -----VREKKEKVM-----KEBEVTAKAPACKADKTLATORRLEERQR 780
QY 856 ELEMAKELKPNEDCLADQKLPPLPRIPGLVLSGSTFSDCLMVVQFLRNFQKVLGFDV 915
Db 781 QQMILEEMKPTEDMCLTDHQPLPDFSRVPGTLPSGAFSDCLTIVEFLHSFGVLGFD 840
QY 916 NIDVNLVLQGLLNI GDSMEVQDILLVRLLSAAVCDPGLITGYKAKTALGEHLLNVG 975
Db 841 AKDVPSLGLVQEGLLCQGDLSGEVQDILLVRLKAAHLHDGFPSPYQCSLKILGEKVSEIPL 900
QY 976 NRDNVSEILQIFMEAHCGQTELTESIKTAKFAHTPAOKASVLAEFLNELACKSVVSEI 1035
Db 901 TRDNVSEILRCFIMAYGEVAPALCDRLRQTPQAPQPPQKAAVLAEFLVHELNGSTLIINEI 960
QY 1036 DKNIDYMSLRDKWVVEGKRLKRIIHAKTGKERTDSTGGIDGEEHQPLGTPPTGPKRR 1095
Db 961 DKTLESMSYRNKVI VEGRLRLKTVLAKTRGSE-----VMEGPEECLG-----RR 1009
QY 1096 RKGGSDDYDDDDSDGDE 1155
Db 1010 RSSRIME-----ETSGMEEEEEEESIAAIVPGRGR--RDGEVDATASSIPBLERQI 1058
QY 1156 EKLKQOOSYRRLKFLDASHLSRSMFGPDYRPRYRWWILPRCGIIFVEGSEGELEE-IA 1214
Db 1059 EKLKSRQFLFRKXLLHSSQMLRAVSLGQDRYRYYWVLPYLAGIFVEGTGNLVPVEEVIK 1118
QY 1215 KERELKKAESVQIKENWF-----ETSGDSLNCNTDHCBEQKEDLKEKNTNLFLOKPGSF 1270
Db 1119 KETDSLKVNAHASINPALFPMKXNELAGSNTTASSPARAGRPR-----KTKPGSM 1169
QY 1271 SKLSKLLEY-----AKMPPESEVMTTP-----KPNAGANGCTLSYQNS 1307

Db 1169 QPRHLKSPVRGQDSQPOAQLOPEAQLOHAPQPOLOQLOLQSHKGFLEQEGSPISLGOS 1228
QY 1308 GKHSLSGVSQS-----TATQSNVEKADSNLNTGSSGKFYSPLPND-----QLLK 1354
Db 1229 -QHDLN--QSAFLSWLSQTSQSHSSLLSSVL--TPDSSFGKL-DRAPSQPPPEPDEAE 1282
QY 1355 TLTEKRNQWFSLLPRTPCDDTSLTHADMSTASLVTPSQSPSSKS---PSPTAPILGSS-- 1409
Db 1283 SSPDQALWFNISQAQPCN-----AAPTp---PPAVSEDOPTPSQQLASSKP 1327
QY 1410 AQNPVGLNPALSPLOVKGVGSMGL---QFCGWTGVVTSNIPFSLVPSLGSGLGSE 1466
Db 1328 MNRPSAANP--CSPVQF--SSTPLAGLAPKRRAGDFG-----EMPOSPTGLGQPK 1373
QY 1467 GNGNSFLTNSVASSSESFPQNEKATSAQPAAVEKVPDPFSPKPIPEEQFGWRII 1526
Db 1374 RRRG-----PPSKF---FKQMEORYLTQTA-----QVPPEMCSGWWIR 1411
QY 1527 DPEDIKALLKVLHLRGIREKALQKQIQKHLDYITQACLNKNDVAIIELNENEENQVTRDI 1586
Db 1412 DPEMLDMLKALHPRGIREKALHKLNRDPLQEVCLRPSADPIFEPRQLPAFO---EG 1468
QY 1587 VENNASVEQAMEMDLVLQOVEDLERRVASASLQVKGWMCPEPASEREDLVYFEHKSPTK 1646
Db 1469 INMSPKKTYETDILAVLQWVEELQVRVIMSDLQIRGWTCPSPDSTREDLAYCEHLSDS- 1527
QY 1647 LCKEHDGFTGEDESSAHALERKSONPLDIAVTRLADLERNIERR----- 1691
Db 1528 ---QEDITWRGRGREG-APORKTNPDLDLAVRLAALAEQNVERRYLREPLWPTHEVULE 1583
QY 1692 -----IEEDIAPGLRVMRRLSARSAAQVALCICQLOKSIAMWEKSTM 1734
Db 1584 KALLSTPNGAPEGTTTETSIYSBITPRIVWRQTLERCRAAQVCLCLGQLERSIAMKSVN 1643
QY 1735 KYCQICRKGNEBELLLCDGCDKCHTYCHRPKLTITPDGOWFCPCACIAKASGOTLKIK 1794
Db 1644 KVTCLVCRKGNDEFLLLCDGCDRCHTYCHRPKMEAVPEGDWFTCTVCLAQ----- 1694
QY 1795 KLHVKGKKTNE---SKGKKVTLTGDEDEDSASTSSSLKRGNKDQKQKM---EENTS 1847
Db 1695 ---QVEGETQKPGFPKQK-----RKSGYSLNFSSEGDRRRRVLLRGRESA 1740
QY 1848 INLSKQESFTSVKPKP-----RDDSKDALCALCSMILTEMETHEDAWPFLPLVNLKLVPGYKK 1903
Db 1741 AGPRVSEGLSPSKERRLSMRNHSDLTFCIILMEMESHDAAPFLEPVNPRVLSGYRR 1800
QY 1904 VIKPMDFTTIREKLSSQOYNLETFDALDVLDFDNCETFDNEDDDSDIGRAGHMRKYFEK 1963
Db 1801 IIKNPMDFSTWRRLRGYTSSEBFAADALLVFDNCQTFNEDDSEVGKAGHMRPFES 1860
QY 1964 KWTDTFK 1970
Db 1861 RWEFIQ 1867
RESULT 10
AAB95554
ID AAB95554 standard; protein; 1727 AA.
XX
AC AAB95554;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human protein sequence SEQ ID NO:18183.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.

XX 28-JUL-2000; 2000EP-00116126.
 XX
 PR 29-JUL-1999; 93JP-00248036.
 PR 27-AUG-1999; 95JP-00300253.
 PR 11-JAN-2000; 2000JP-00118776.
 PR 02-MAY-2000; 2000JP-00183767.
 PR 09-JUN-2000; 2000JP-00241899.
 XX (HELI-) HELIX RES INST.
 XX
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 XX WPI; 2001-318749/34.
 DR
 XX
 PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-
 PT length cDNAs defined in the specification, and for the detection and/or
 PT diagnosis of the abnormality of the proteins encoded by the full-length
 PT cDNAs.
 PT
 XX
 PS Claim 8; SEQ ID NO 18183; 2537pp + Sequence Listing; English.
 XX
 CC The present invention describes primer sets for synthesizing 5602 full-
 CC length cDNAs defined in the specification. Where a primer set comprises:
 CC (a) an oligo-dT primer and an oligonucleotide complementary to the
 CC complementary strand of a polynucleotide which comprises one of the 5602
 CC nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination
 CC the 5'-end sequence/3'-end sequence is selected from those defined in the
 CC specification. The primer sets can be used in antisense therapy and in
 CC gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893
 CC represent human amino acid sequences; and AAH13629 to AAH13632 represent
 CC oligonucleotides, all of which are used in the exemplification of the
 CC present invention
 XX
 SQ Sequence 1727 AA;
 Query Match 24.6%; Score 2498; DB 4; Length 1727;
 Best Local Similarity 33.8%; Pred. No. 4.7e-155;
 Matches 646; Conservative 297; Mismatches 589; Indels 380; Gaps 56;
 QY 243 PSQKKQSSSKLKVIAALSNPATSSPAHPKQTLNHNHPFLTN--ALLGN----- 295
 DB 6 PSFPTSQTSPLMLGSIQTF-PSQEVSGIHDPDEAEKEMTSVVAENGTLVGSLELEE 64
 QY 286 HOP-----NG---VIQSVIQEAPL-----ALTTKRMQSKINEN--IAASSTPF 335
 DB 65 EQPELMKMGYNSVPSVESLHQEVSVPDPTVSCLDPPSHLPDQLEDTPILSSEDSLEFP 124
 QY 336 SSPVNLSTSGRRTPGNQTPVMPSPILHSQCKEKAQVSNVNPVKT-----QHHSHPAKS 390
 DB 125 NLAPEPVSGGLYGIDDTLM-----GAEDKLPLEDSFVLSALDCPSLNNATAPS 174
 QY 391 LVEQFRGTDSDTPSSKQSDNSNEDDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDE 450
 DB 175 LLADDSQTSISIFASPTSPVLGSEVLQDNSFDLNGNSDAEQEEMETQSDPPPPSLTQPA 234
 QY 451 EEDDDDDKDDSDSDTEGEKTSMLKNTTS----- 481
 DB 235 -----PDQSSTIQLHPATSPAVSPPTSPAVSLVSPSPASPSISPEVCPAASVTVSP 285
 QY 482 ---SVKSPSMS-----LTGHSTPRNLHIKAPGSA--PAALCSESQSAPFLGTSS 526

DB AVFSWSPASSAVLPAVSLEVLPTASVTSPKASPVTSAAAAPTASPANKDVDSFLETTA 345
 QY 527 ST-----LTSSPHSGTSKRRRYTDERELRI PLEYGQRETRIRNFGRLQGEVAYYAP 579
 DB 346 DVEBITGEGLTAS--GSGDVMRRCIATPEEVLRLPQHGMRRREVRIKKGSHRWQGETWYVGP 404
 QY 580 CGKKLROYPEVIKYLSRNGIMDISRDNFSFAKTRVGDIFYEARDGPOEMOMCLKEEDVI 639
 DB 405 CGKMKQFPPEVIKYLSRNVHSVREHFSRPMVGDFFFEERTDTPGLQHWQVLSAEIIP 464
 QY 640 PRIMEGRGRPPNPDRORAREESRMRKGRPPNVGNAAFFLDNADAKLRLKLOAQSI 699
 DB 465 SRIQAITGKGRPRNTEKAKTKEVPKVRGRGRPPKVKITELLNKTNRPLKLEAQB-- 522
 QY 700 RQAAQIKLLRKLQOQARVAKEAKQ-----QAIMAAEEKRRKQKQIKMKQOEK 750
 DB 523 -----TLNEDKAKIAKSKKVRQKVRQGECCOTTIOGQARNKRKQTKSLKQK-- 570
 QY 751 IKRIQIRMEKELRAQOILEAKKKKEEANAANAKLEAKRIKEKEMRRQQAVALLKQBRE 810
 DB 571 -----EAKKSKAKKEGK--TKQEKLEK-----VKREKKE 600
 QY 811 RRRQHMLMKAMEARKKAEERLQKBRDEKRLNKRERLQKLEORLELEMAKELKKPNEDM 870
 DB 601 KVKM-----KEKEEVTKAKPACKADKTLATQRLLEERQKQOMILEEMKKPTEDM 649
 QY 871 CLADQKLPPELPRIPLGLVSGSTSDCLMVQVFNFGKVLGFDVNDVNLVLOEGLL 930
 DB 650 CLTDHQPLPDFSRVPGTLPFGAFSDCLTIVFELHSGFVKVLGFPDPAKDVSLGVLQEGLL 709
 QY 931 NIGDSMEVODLLVRLLSAAVCDPLITGVKAKTALGEHLNVGNVNDVNSEIILQIFMEA 990
 DB 710 CQGSLEGEVQDLLVRLKAAHLHDGFPFSCOSLKILEKVSSEIPLTRDENSEILRCFLMA 769
 QY 991 HCGOTELTESLTKTAKQAHTPAQKASVLAFLINELACSKSVSVSEIDKNIDYMSNLRDVK 1050
 DB 770 YGVEPALCDRLRTQPFQAQPPQQAQVLAFLVHLNGSTLIINEIDKTLSEMSVSKYKNKW 829
 QY 1051 VVEGKLRLRIHAKTKGRTSGDILGEBQHPLTPTPGKRRKRGDSDYDDDDDD 1110
 DB 830 IVEGRLRLKTVLAKRTGRSE-----VMEGPEECIG-----RRRSRIME----- 870
 QY 1111 SDDGQDEDEDEDEKQKGTIDICDEDEGDQAAASVEELEKQIEKLSKQSOVRRKLF 1170
 DB 871 -ETSGMEEEESIAAVPGRGR--RDGEVDATASSIPELERQIEKLSKQLFRKLL 927
 QY 1171 DASHLSRVMPGPDYRRRYWILPRCGGIFVEGMEGEGLEE--IAKEREKLLKKAESVQIK 1229
 DB 928 HSSQMLRAVSLGQDRYRRRYWVLPYLAGIFVEGTEGNLVPEVIKKTETSLKVAHAASLN 987
 QY 1230 EEMF-----ETSGDSLNCSTHCEQKEDLKEKNTNLFLOKPGSFSLKSLLEV----- 1279
 DB 988 PALFSMKMELAGSNTTASSPARARGRP-----KTKPGSQMPRHLSKSPVRGQDSE 1037
 QY 1280 ---AKMPESSEVMT-----KPNAGANGCTLSYQNSGKHSLSGVOS----- 1317
 DB 1038 QPQQLQPEALHAPQAPQPOLQLOLQSHKGLFLEQESPLSLGQS--QHDLS--QSAFLSW 1094
 QY 1318 -TATQSNVVEKADSNLNFNTGSGPGKFSYPLPND-----QLLKTILTEKRNQWPSLLPR 1369
 DB 1095 LSQTQSHSLSSSVL--TPDSSPGKL-DPAPSQPPEPEPEDEAESPPQALWFNISAQ 1151
 QY 1370 TPCDDTLTHADMTASLVTQSQPPSKS----PSPTAPLQSS--AQNPVGNIPALSPL 1424
 DB 1152 MPCN-----AAPT-----PPAVSEDOPTSPQOLASSKPMNRPSAAN--CSGPV 1194
 QY 1425 QVKGVSNMGL---QFCGWPTGVVTSNIPFTLSVPSLGSGLGSEGNNSFLTNSVASSK 1481
 DB 1195 QF-SSTPLAGLAPKRRAGDPG-----EMQSPDTGLGQPKRGR-----PFSK 1235
 QY 1482 SESVPQNEKATSAQAAVEVAKPVPFPSPKPIPEEMQFGWRIIDPEDLKALLKVLHLR 1541

Db 1236 F---FKQMEQRVLTQTA-----QVPPMCSGWWRDPEMLDKALHPR 1280
 Qy 1542 GIREKALQKQI QKLDYITQACLKKNQDAI IELNENEENQVTRDIVENWVSEQAMMDL 1601
 Db 1281 GIREKALHKLHNRDFLQEVCLRPADP IEPRLQPAFQ---EGIMSPKREKTYETDL 1337
 Qy 1602 SVLQOVEDLERRVASLQVGMKCPPEASEREDLVYFHKHFTKLCCKHDEGFTGDEBS 1661
 Db 1338 AVLQWVEELEQRVIMSDLQIRGWTCPSPDSTREDLAYCEHLSDS---QEDITWRGORE 1393
 Qy 1662 SAHALERKSDNPLDIAVTLADLERNIERR----- 1691
 Db 1394 GL-APQKTTNPLDLAVMLAALAEQNVERRYLRPLWPHTEVLEKALLSTNGAPGTT 1452
 Qy 1692 --IEEDIAPLGLRWRRALSEARSAAQVALCQLOKSI AWEKSIIMKYQICIRKGDNEBL 1749
 Db 1453 TEISYEITPRIRVWRQTLERCSRSAQVCLCQLGERSIAWEKSVNKVTCVCRKGDNEF 1512
 Qy 1750 LLLCGDCXGCHYCHRPKITIPDGDWFCPCACI AKASQTLKIKLHVKGKKTNE---S 1806
 Db 1513 LLLCGDCRGCYCHRPKMEAVPEGDWFCVCLAQ-----QVEGEFTQKPGFP 1561
 Qy 1807 KKGKVTTLTGDTEDSDASTSSLRGNKDLQKRM---EENTSNLSKQESFTSVKXP 1862
 Db 1562 KRGQK-----RKSGYSLNPFSEGDGRRRRVLLRGRESPAAGPRYSEGLSPSKR 1609
 Qy 1863 K---RDDS KOLALCSMILTEMETHEDAWPFLLPVNLKLVPGYKVKIKKPMDFSTIREKL 1918
 Db 1610 RLRLSMNHHSDLTFCIEILLWEMESHDAWPFLEPNPLVSGYRIIKNPMDFTWREBL 1669
 Qy 1919 SSGQVPLNLTFFALDVLVLPDNCETNEDDSDIGRAGHNRKRYFEKKWTDTFK 1970
 Db 1670 LRGGYTSSEFAADALLVDPNCQTFNEDDSEVGKAGHMRFRFESRWEFFYQ 1721

RESULT 11

AAW81170
 ID AAW81170 standard; protein; 1878 AA.
 AC AAW81170;
 XX
 DT 05-MAY-1999 (first entry)
 XX
 DE Human BAZ2-alpha protein.
 XX
 KW Transcriptional regulator; transcription; BAZ1-alpha; bromodomain; BAZ;
 KW atypical zinc finger; testis; human; tumour; BAZ1-beta; BAZ2-alpha; drug;
 KW BAZ2-beta; treatment; cancer; proliferative disorder; screening.
 XX
 OS Homo sapiens.
 XX
 PN WO9847920-A1.
 XX
 PD 29-OCT-1998.
 XX
 PF 17-APR-1998; 98WO-JP001783.
 XX
 PR 18-APR-1997; 97JP-00116570.
 PR 24-OCT-1997; 97JP-00310027.
 XX
 PA (CHUS) CHUGAI RES INST MOLECULAR MEDICINE INC.
 XX
 PI Jones MH;
 XX
 DR WPI: 1998-583603/49.
 DR N-PSDB; AAV68396.
 XX
 PT Transcriptional regulator gene family containing bromodomain - may be
 PT expressed in testis tissue and is useful for treatment of cancer and
 PT other proliferative disorders.
 XX
 PS Claim 1; Page 72-88; 187pp; Japanese.

CC This sequence represents the human BAZ2-alpha protein, a member of a
 CC family of transcriptional regulator genes containing a bromodomain (BAZ,
 CC Bromodomain with Atypical Zinc finger) which are expressed specifically
 CC in testis tissue and also in certain tumour lines. Transgenic cells may
 CC be used for the preparation of the BAZ1-alpha, BAZ1-beta, BAZ2-alpha and
 CC BAZ2-beta proteins. These proteins can be used in the treatment of cancer
 CC and other proliferative disorders, and in screening of compounds for
 CC their binding ability to the expression products (e.g. for use as drugs
 CC by modulation of transcriptional regulation)
 XX
 SQ Sequence 1878 AA;

Query Match 24.6%; Score 2493.5; DB 2; Length 1878;
 Best Local Similarity 32.2%; Pred. No. 1.1e-154;
 Matches 681; Conservative 331; Mismatches 675; Indels 426; Gaps 68;

Qy 43 MESSNSDSD----SGTSSDTSSEGISDSDLDLEDEEEDDSDSESEAO 98
 Db 1 MEMEANEANDHNFNFTGLPPAPAAAGLKPSPS-----SDEGLYTNNGSPNFPQ 48
 Qy 99 HKS-NNQVLLHG---ISDPKADG--QKATERAQEKRIHQPLPLAPE-----SOTHSFQ-- 145
 Db 49 CKSLNGDVNAGLSTVSTHTTSGILNSAPHSSTSHLHP-SVAYDCLWVSYQFSANPG 107
 Qy 146 SOQKQPVLSQOLPPIFOSSQAKEESVNNKHTSVIQTGLVSNVPLSLVNAQAKKETYMKL 205
 Db 108 SNLKDPPLLSQ-----FSGQVPLNGI-----LGGSRQPS----- 138
 Qy 206 IVPSPDV-LKAGNKVTSSESLTSELASKREQYKQAPPSQLKQESSKSLKVTAALSN 264
 Db 139 --PSHNTWLRAGSQKFWANGTHSPMGLNFDQELYSFPDQ-----NFEVCSGI-- 186
 Qy 265 PKATSSSPAHPKQTLNHNHPFLTNAL-----LGNHQP-----NG---VTSQVIQ 307
 Db 187 -----HPEAAEKEMTSVVAENGTLVCSLEEKQELKMCNGVSGVPSVESHQ 237
 Qy 308 EAPL-----ALTTKTMOSKINEN--IAAASSTPFSSPVNLTSGRRTPGNQTPVP 357
 Db 238 EVSLVPDPTVSCLDPPSHLPDQLEDTPILSEDSLEPPNLSLAPESPVSGLYGIDDTLM- 296
 Qy 358 SASPILHSQGEKAVNNVNPVKT-----QHHSHPAKSLVEQFGRTDSDIPSSKQSESN 412
 Db 297 -----GAEDKLPLEDSEFVSALDCPSLNNATAFSLADDSQTSITSIPASPTSPVL 347
 Qy 413 EDEEDDEEEDDEDDSDSDSESNSSESDETEGSEEDDDDDKQDESDSDTEGKT 472
 Db 348 GESVLQDINSFDLNGSDAEQEMETQSSDFPSPUTOPA-----PDQSTIQLHPAT 398
 Qy 473 SMKLNKTTT-----SVKSPSMS----- 489
 Db 399 SPAVSPITSPAVSLVSPAPASPEISPEVCPAASTVSPVSPVSPASSAVLPVSLVLP 458
 Qy 490 LTGHSTPNLHIAKAPGA-PAALCSESQSPAFIQTSSST-----LTSPHSGTSKRR 541
 Db 459 LTASVTPSPKSPVTPSPAAAFTAGSPANKDVSSFLTTADVEEITGEGLTAS-GSGDVMR 517
 Qy 542 RVTDERELRIPLEYGWORETRIRNFGGRLOGEVAYAPCGKKLRQYBEVIKYLNRGMD 601
 Db 518 RIATPEEVRLPLQHGWRREVRIRKKGSHRWQGETWYGPCGRMKQFPBEVIKYLNRVHS 577
 Qy 602 ISRDNFSFSAXIRVGDIFYEARDGPQEMQWCLLKEDVPIPRIRAMEGRGRPPNPDQRAR 661
 Db 578 VRREHFSFSPMPVGDFFERDTPGLQWVLSAEIIPSRIOAITGKGRPRNTEKATK 637
 Qy 662 EESRMRKGRPPNVGNABFLDNADAKLLKLOAEIARQAQIKLLKLOKQEQARVAK 721
 Db 638 EVPKVRGRGRPPKVKITELNKTNDNRPLKLEAQ-----TLNEEDKAKIAK 685
 Qy 722 EAKQQAQIAAEKREKQEQIKIMKQEKIKRIQIIRMEKELRAQQLILEAKKKKEAAN 781
 Db 686 SKKKMR-----QKVQGECLTTIQQARNRKQETKSLKH-----KEAKKKKAEKEK 733
 Qy 782 AKLLEAEKRIKEKEMRRQAVLLKHQERRRRQHMLMLKAMEARKKAEKERLQEKRDE 841

utilisation of the activities such as: Immune system suppression, Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic and thrombolytic activity, cancer diagnosis and therapy, drug screening, assays for receptor activity, arthritis and inflammation, leukaemias and C.N.S disorders. Note: The sequence data for this patent did not form part of the printed specification

Sequence 1589 AA;

Query Match 24.6%; Score 2492; DB 4; Length 1589;
Best Local Similarity 35.9%; Pred. No. 1e-154;
Matches 611; Conservative 260; Mismatches 492; Indels 338; Gaps 48;

431	QY	ESDSSQSDNSGSDSTEGSEBDDDKD-----QDESODTGEKTSMKLNKTTTS	481
18	Db	ESVLQDNSFDLNGSDAE-QEEMETOSSDPFSLTQAPDQSSSTIQLHPATSPAVSPPTS	76
482	QY	-----SVKSPSMS-----LICHSTPRN	498
77	Db	PAVSLVVSIPAASPEISPEVCPAASTVVSFAVSVSPASSAVLPAVSLEVLPTASVTSPK	136
499	QY	LHIAKAPGSA-PAALCSSESOPAFLTSSST-----LTSSHSIGTSKERRVTDREL	550
137	Db	ASPVTSFAAAPTASPAKNDVVSFLETTADVEHITGEGLTAS-GSGDWMRRRTATPEVR	195
551	QY	IPLYQWQRETRIRNFGRLQGEVAYAPCGKKLRQYPEVIKYLRSRNGIMDISRDNFSFS	610
196	Db	LPLQHGWRREVRIGKSHRWQGETWYVPCGKRMQFPEVIKYLRSNVHVSRRHEHFSFS	255
611	QY	AKIRVGDFFEARQGPQEMQWCLLKEDVIPRIAMEGRRGRPPNDPRQRAEESRMRRK	670
256	Db	PRMPVGDFPEERDTPEGLQWVLSAEIPSIQAITGKRGRPRNTEKARTKEVPKVRGR	315
671	QY	GRPPNVGNAEFLDNADAKLLRKLQAOEIAQAAQIKLLRKLQOEQARVAKEAKQ-----	726
316	Db	GRPPKVYITELLNKTNDNRPLKLEAQB-----TLNEDKAKIAKKKKRQKV	363
727	QY	-----QAIMAAEKKRQKEQIKIMQOEKIKRIQOIRMEKELAAQQLILEAKKKKEBAAN	781
364	Db	QRGECQTTIQOARNKRKQETSLKQK-----EAKKSKAEKEK	402
782	QY	AKLLEAKKIKEKEMRRQOAVLLKHQERERRRQHMLMKAMARKAAEAKERLQOEKXDE	841
403	Db	GK--TKOEKLKEX-----VKREKKEKVKM-----KEKEEYTKAPACKAD	440
842	QY	KRLNKRKLEORLELEMAKELKPNEDMCLADQKPLPELPRIPGLVLSGTSFSDCLMVV	901
441	Db	KTLATORLEERQRMILEEMKKPTEDMCLTDHQPDPFSRFPGLPLPSGAFSDCLTIV	500
902	QY	QFLRNFKGVLGFDVNIDVNLVSLQBELGLNIGDSMGVEQDLLVRLLSAAVCDPGLITGYK	961
501	Db	EFLHSFGVLGFDPAKDVPSLGLVQSGLLCOGDSLGEVDLLVRLKAAHLDFGFPSCYQ	560
962	QY	AKTALGEHLLNVGNRDNVSEILQIPWEAHCQGTETTESLTKFAQHTPAQKASVLAF	1021
561	Db	SLKILGEKVSEIPLTRDNVSEILRCFLMAYGVEPALCDRLRTPFOAQPPQKAAVLAF	620
1022	QY	INELACKSVVSEIDKNIDYMSNLRDQKVVGEKLRKLRIIHAKKTKGRTDSGGIDLEE	1081
621	Db	VHELNGSTLIINBIDKLTSMSSYRKNKVIVEGRRLRLKTVLAKRTGRSE-----VMEVG	676
1082	QY	QHPLGTPTPKRRRRKGGSDYDDDDDDDDDEDEDEDEKEDQKGGKTTDICEDEDE	1141
677	Db	ESCLG-----RRSSRINEV-----TSGMEEEEEEESIAAVPGRGR--RDCEV	718
1142	QY	GBOAASVELEKQIEKLSKQSQSYRRKLFDAHSLSRWSMFGPDYRRRYWILPRCGGIFV	1201
719	Db	DATASSIPELERQIEKLSKQLPFRKLLHSSQWLRVAVSLGQDRYRRRYWVLPVLAGIFV	778
1202	QY	EGMESGEGLEE-IAKREKLIKAEBSVQIKEMF-----ETSGDSLNCNTDHCQKEDLKE	1256
779	Db	ECTEGLNVPEEVIKETSDLSKVAHAASLNPALFMSKMWELAGSNTTASSPARAGRPR--	835

KW Human; foetal liver; gene expression; single exon nucleic acid probe.
XX Homo sapiens.
XX WO200157277-A2.
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US000669.
XX
XX 04-FEB-2000; 2000US-0180312P.
XX 26-MAY-2000; 2000US-0207456P.
XX 30-JUN-2000; 2000US-00608408.
XX 03-AUG-2000; 2000US-00632366.
XX 21-SEP-2000; 2000US-0234687P.
XX 27-SEP-2000; 2000US-0236359P.
XX 04-OCT-2000; 2000GB-00024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
PI WPI; 2001-483447/52.
XX Human genome-derived single exon nucleic acid probes useful for analyzing
PT gene expression in human foetal liver.
XX
XX Claim 27; SEQ ID NO 25672; 639pp + Sequence Listing; English.
XX The invention relates to a single exon nucleic acid probe for measuring
CC human gene expression in a sample derived from human foetal liver. The
CC single exon nucleic acid probes may be used for predicting, measuring and
CC displaying gene expression in samples derived from human foetal liver. The
CC present sequence is a peptide encoded by a single exon nucleic acid probe
CC of the invention. Note: The sequence data for this patent did not form
CC part of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 194 AA;
SQ
Query Match 10.1%; Score 1021; DB 4; Length 194;
Best Local Similarity 100.0%; Pred. No. 5.3e-59;
Matches 194; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1232 MFETSGDSLNCNTDHCQEKEDLKEKDNNTNLFQKPGSFSLKLELVAKMPPESEVMT 1291
Db 1 MFETSGDSLNCNTDHCQEKEDLKEKDNNTNLFQKPGSFSLKLELVAKMPPESEVMT 60
QY 1292 KPNAGANGCTLSYQNSGKHSLSGVOSTATQSNVEKADSNLNTGSSGPGKFYSPLPNDQ 1351
Db 61 KPNAGANGCTLSYQNSGKHSLSGVOSTATQSNVEKADSNLNTGSSGPGKFYSPLPNDQ 120
QY 1352 LLKLTITEKNROWFSLPRTPCDDTSLTHADMSTASLVTPQSPKSPSPPTAPLGSSAQ 1411
Db 121 LLKLTITEKNROWFSLPRTPCDDTSLTHADMSTASLVTPQSPKSPSPPTAPLGSSAQ 180
QY 1412 NPVGLNPPFALSPLQ 1425
Db 181 NPVGLNPPFALSPLQ 194
RESULT 18
AAM26499
ID AAM26499 standard; protein; 194 AA.
XX
XX AAM26499;
XX
XX 17-OCT-2001 (first entry)
XX Peptide #536 encoded by probe for measuring placental gene expression.
XX Probe; microarray; human; placenta; antenatal diagnosis;
KW genetic disorder.

XX Homo sapiens.
XX WO200157272-A2.
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US000663.
XX
XX 04-FEB-2000; 2000US-0180312P.
XX 26-MAY-2000; 2000US-0207456P.
XX 30-JUN-2000; 2000US-00608408.
XX 03-AUG-2000; 2000US-00632366.
XX 21-SEP-2000; 2000US-0234687P.
XX 27-SEP-2000; 2000US-0236359P.
XX 04-OCT-2000; 2000GB-00024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
PI WPI; 2001-488897/53.
XX Human genome-derived single exon nucleic acid probes useful for analyzing
PT gene expression in human placenta.
XX
XX Claim 27; SEQ ID NO 26768; 654pp; English.
XX The present invention relates to single exon nucleic acid probes (SENP;
CC see AA31315-A157546). The present sequence is a peptide encoded by one
CC such probe. The probes are useful for producing a microarray for
CC predicting, measuring and displaying gene expression in samples derived
CC from human placenta. The probes are useful for antenatal diagnosis of
CC human genetic disorders
XX
XX Sequence 194 AA;
SQ
Query Match 10.1%; Score 1021; DB 4; Length 194;
Best Local Similarity 100.0%; Pred. No. 5.3e-59;
Matches 194; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1232 MFETSGDSLNCNTDHCQEKEDLKEKDNNTNLFQKPGSFSLKLELVAKMPPESEVMT 1291
Db 1 MFETSGDSLNCNTDHCQEKEDLKEKDNNTNLFQKPGSFSLKLELVAKMPPESEVMT 60
QY 1292 KPNAGANGCTLSYQNSGKHSLSGVOSTATQSNVEKADSNLNTGSSGPGKFYSPLPNDQ 1351
Db 61 KPNAGANGCTLSYQNSGKHSLSGVOSTATQSNVEKADSNLNTGSSGPGKFYSPLPNDQ 120
QY 1352 LLKLTITEKNROWFSLPRTPCDDTSLTHADMSTASLVTPQSPKSPSPPTAPLGSSAQ 1411
Db 121 LLKLTITEKNROWFSLPRTPCDDTSLTHADMSTASLVTPQSPKSPSPPTAPLGSSAQ 180
QY 1412 NPVGLNPPFALSPLQ 1425
Db 181 NPVGLNPPFALSPLQ 194
RESULT 19
ABB27866
ID ABB27866 standard; peptide; 194 AA.
XX
XX ABB27866;
XX
XX 01-FEB-2002 (first entry)
XX Human peptide #517 encoded by breast cell single exon nucleic acid probe.
XX Human; microarray; single exon probe; gene expression; breast; disease;
XX cancer.
XX Homo sapiens.
XX

PN WO200157271-A2.
XX
PD
XX
XX 09-AUG-2001.
PF 30-JAN-2001; 2001WO-US000662.
XX
XX 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-496933/54.
XX
PT New spatially-addressable set of single exon nucleic acid probes, useful
PT for measuring gene expression in sample derived from human breast,
PT comprises number of single exon nucleic acid probes.
XX
XX Claim 27; SEQ ID NO 10834; 327pp + Sequence Listing; English.
XX
XX The invention relates to a spatially-addressable set of single exon
XX nucleic acid probes for measuring gene expression in a sample derived
XX from human breast and BT 474 cells. The method involves contacting the
XX probes with a collection of detectably labelled nucleic acids derived
XX from mRNA of human breast, and then measuring the label bound to each
XX probe of the microarray. The probes are useful for verifying the
XX expression of regions of genomic DNA predicted to encode proteins. They
XX are useful for gene discovery, and for determining predisposition and/or
XX prognosing breast disease. Gene expression analysis is useful for
XX assessing the toxicity of chemical agents on cells. The microarray of
XX this invention presents a far greater diversity of probes for measuring
XX gene expression, with far less bias than expressed sequence tag
XX microarrays. The method is suitable for rapid production of functional
XX information from genomic sequence. The present sequence is a peptide
XX encoded by a single exon nucleic acid probe of the invention. Note: The
XX sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 194 AA;
Query Match 10.1%; Score 1021; DB 4; Length 194;
Best Local Similarity 100.0%; Pred. No. 5.3e-59;
Matches 194; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1232 MFETSGDSLNCSTHDCQKEDLKEKNTNLFLOKPGFSKLSKLLVAKMPPSEVMT 1291
Db 1 MFETSGDSLNCSTHDCQKEDLKEKNTNLFLOKPGFSKLSKLLVAKMPPSEVMT 60
QY 1292 KPNAGANGCTLSYQNSGKHSILGVSQSTATQSNVEKADSNLNTGSSGPGKFYSLPNDQ 1351
Db 61 KPNAGANGCTLSYQNSGKHSILGVSQSTATQSNVEKADSNLNTGSSGPGKFYSLPNDQ 120
QY 1352 LLKLTTEKNROWFSLLPRTPCDDTSLTHADNSTASLTVPQSQPPSKSPSPAPLGSSAQ 1411
Db 121 LLKLTTEKNROWFSLLPRTPCDDTSLTHADNSTASLTVPQSQPPSKSPSPAPLGSSAQ 180
QY 1412 NPVGLNPFALSPQ 1425
Db 181 NPVGLNPFALSPQ 194
RESULT 20
ID ABB18510
XX ABB18510 standard; protein; 194 AA.
XX
AC ABB18510;

XX 23-JAN-2002 (first entry)
XX
XX Protein #509 encoded by probe for measuring heart cell gene expression.
XX
XX Human; gene expression; heart; microarray; vascular system;
XX cardiovascular disease; hypertension; cardiac arrhythmia;
XX congenital heart disease.
XX
XX Homo sapiens.
XX
XX WO200157274-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US000666.
XX
XX 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-488899/53.
XX
XX Single exon nucleic acid probes for analyzing gene expression in human
XX hearts.
XX
XX Claim 15; SEQ ID NO 20280; 530pp; English.
XX
XX The present invention relates to single exon nucleic acid probes for
XX measuring human gene expression in a sample derived from human heart (see
XX AB21335-AB41305). The present sequence is a protein encoded by one such
XX probe. The probes may be used for predicting, measuring and displaying
XX gene expression in samples derived from the human heart via microarrays.
XX By measuring gene expression, the probes are useful for predicting,
XX diagnosing, grading, staging, monitoring and prognosing diseases of the
XX human heart and vascular system e.g. cardiovascular disease.
XX hypertension, cardiac arrhythmias and congenital heart disease. Note: The
XX sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 194 AA;
Query Match 10.1%; Score 1021; DB 4; Length 194;
Best Local Similarity 100.0%; Pred. No. 5.3e-59;
Matches 194; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1232 MFETSGDSLNCSTHDCQKEDLKEKNTNLFLOKPGFSKLSKLLVAKMPPSEVMT 1291
Db 1 MFETSGDSLNCSTHDCQKEDLKEKNTNLFLOKPGFSKLSKLLVAKMPPSEVMT 60
QY 1292 KPNAGANGCTLSYQNSGKHSILGVSQSTATQSNVEKADSNLNTGSSGPGKFYSLPNDQ 1351
Db 61 KPNAGANGCTLSYQNSGKHSILGVSQSTATQSNVEKADSNLNTGSSGPGKFYSLPNDQ 120
QY 1352 LLKLTTEKNROWFSLLPRTPCDDTSLTHADNSTASLTVPQSQPPSKSPSPAPLGSSAQ 1411
Db 121 LLKLTTEKNROWFSLLPRTPCDDTSLTHADNSTASLTVPQSQPPSKSPSPAPLGSSAQ 180
QY 1412 NPVGLNPFALSPQ 1425
Db 181 NPVGLNPFALSPQ 194
RESULT 21

AAM66222
ID AAM66222 standard; protein; 194 AA.
XX AC
XX AAM66222;
XX AC
XX 06-NOV-2001 (first entry)
XX DT
XX XX
XX Human bone marrow expressed probe encoded protein SEQ ID NO: 26528.
DE Human; bone marrow expressed exon; gene expression analysis; probe;
KW Human; bone marrow expressed exon; gene expression analysis; probe;
KW microarray; cancer; leukaemia; lymphoma; myeloma.
XX OS
XX Homo sapiens.
XX PN
XX WO200157276-A2.
XX PD
XX 09-AUG-2001.
XX PF
XX 30-JAN-2001; 2001WO-US000668.
XX PR
XX 04-FEB-2000; 2000US-0180312P.
XX PR
XX 26-MAY-2000; 2000US-0207456P.
XX PR
XX 30-JUN-2000; 2000US-00608408.
XX PR
XX 03-AUG-2000; 2000US-00632366.
XX PR
XX 21-SEP-2000; 2000US-0234687P.
XX PR
XX 27-SEP-2000; 2000US-0236359P.
XX PR
XX 04-OCT-2000; 2000GB-00024263.
XX PA
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX PI
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX DR
XX WPI; 2001-488900/53.
XX PT
XX Human genome-derived single exon nucleic acid probes useful for analyzing
XX gene expression in human bone marrow.
XX PS
XX Example 4; SEQ ID NO 26528; 658pp + Sequence Listing; English.
XX CC
XX The present invention provides a number of single exon nucleic acid
XX probes which are derived from genomic sequences expressed in the human
XX bone marrow. They can be used to measure gene expression in bone marrow
XX samples, which may enable the improved diagnosis and treatment of cancers
XX such as lymphoma, leukaemia and myeloma. The present sequence is a
XX protein encoded by one of the probes of the invention
XX SQ
XX Sequence 194 AA;
Query Match 10.1%; Score 1021; DB 4; Length 194;
Best Local Similarity 100.0%; Pred. No. 5.3e-59;
Matches 194; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1232 MFETSGDSLNCSTHCEQKEDLKEKDNLTNLFLOKPGSFSLKLLLEKVPPESEVMT 1291
DB 1 MFETSGDSLNCSTHCEQKEDLKEKDNLTNLFLOKPGSFSLKLLLEKVPPESEVMT 60
QY 1292 KPNAGANGCTLSYQNSGKHSLGVSQSTATQSNVEKADSNLNTGSSGPGKFSPLPNDQ 1351
DB 61 KPNAGANGCTLSYQNSGKHSLGVSQSTATQSNVEKADSNLNTGSSGPGKFSPLPNDQ 120
QY 1352 LKLTLEKRNQWFSLLRTPCDDTSLTHADMSTASLVTPQSQPPSKSPPTPAPLGSSAQ 1411
DB 121 LKLTLEKRNQWFSLLRTPCDDTSLTHADMSTASLVTPQSQPPSKSPPTPAPLGSSAQ 180
QY 1412 NPVGLNPFALSPQ 1425
DB 181 NPVGLNPFALSPQ 194
RESULT 22
AAM53836
ID AAM53836 standard; protein; 194 AA.
XX AC
XX AAM53836;
XX AC
XX 05-NOV-2001 (first entry)
XX DE
XX Human brain expressed single exon probe encoded protein SEQ ID NO: 25941.
XX KW
XX Human; brain expressed exon; gene expression analysis; probe; microarray;
KW Alzheimer's disease; multiple sclerosis; schizophrenia; epilepsy; cancer.
XX OS
XX Homo sapiens.
XX PN
XX WO200157275-A2.
XX PD
XX 09-AUG-2001.
XX PF
XX 30-JAN-2001; 2001WO-US000667.
XX PR
XX 04-FEB-2000; 2000US-0180312P.
XX PR
XX 26-MAY-2000; 2000US-0207456P.
XX PR
XX 30-JUN-2000; 2000US-00608408.
XX PR
XX 03-AUG-2000; 2000US-00632366.
XX PR
XX 21-SEP-2000; 2000US-0234687P.
XX PR
XX 27-SEP-2000; 2000US-0236359P.
XX PR
XX 04-OCT-2000; 2000GB-00024263.
XX PA
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX PI
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX DR
XX WPI; 2001-483446/52.
XX PT
XX Single exon nucleic acid probes for analyzing gene expression in human
XX brains.
XX PS
XX Example 4; SEQ ID NO 25941; 650pp + Sequence Listing; English.
XX CC
XX The present invention provides a number of single exon nucleic acid
XX probes which are derived from genomic sequences expressed in the human
XX brain. They can be used to measure gene expression in brain cell samples,
XX which may enable the diagnosis and improved treatment of nervous system
XX diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia, f
XX epilepsy and cancer. The present sequence is a protein encoded by one of
XX the probes of the invention
XX SQ
XX Sequence 194 AA;
Query Match 10.1%; Score 1021; DB 4; Length 194;
Best Local Similarity 100.0%; Pred. No. 5.3e-59;
Matches 194; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1232 MFETSGDSLNCSTHCEQKEDLKEKDNLTNLFLOKPGSFSLKLLLEKVPPESEVMT 1291
DB 1 MFETSGDSLNCSTHCEQKEDLKEKDNLTNLFLOKPGSFSLKLLLEKVPPESEVMT 60
QY 1292 KPNAGANGCTLSYQNSGKHSLGVSQSTATQSNVEKADSNLNTGSSGPGKFSPLPNDQ 1351
DB 61 KPNAGANGCTLSYQNSGKHSLGVSQSTATQSNVEKADSNLNTGSSGPGKFSPLPNDQ 120
QY 1352 LKLTLEKRNQWFSLLRTPCDDTSLTHADMSTASLVTPQSQPPSKSPPTPAPLGSSAQ 1411
DB 121 LKLTLEKRNQWFSLLRTPCDDTSLTHADMSTASLVTPQSQPPSKSPPTPAPLGSSAQ 180
QY 1412 NPVGLNPFALSPQ 1425
DB 181 NPVGLNPFALSPQ 194
RESULT 23
ABG47890
ID ABG47890 standard; peptide; 194 AA.
XX AC
XX ABG47890;
XX AC
XX ABG47890;

DT 25-FEB-2003 (first entry)
XX Human liver peptide, SEQ ID NO 26538.
XX
XX Human; liver; cirrhosis; hyperlipoproteinaemia; hyperlipidaemia;
KW hypercholesterolaemia; coronary heart disease.
XX
XX Homo sapiens.
XX
XX WO200157273-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US000664.
XX
XX 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-48898/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for analyzing
PT gene expression in human adult liver.
XX
XX Claim 27; SEQ ID NO 26538; 658pp; English.
XX
XX The invention relates to a single exon nucleic acid probe (SENp) (I) for
CC measuring human gene expression in a sample derived from human adult
CC liver, comprising one of 13109 defined nucleotide sequences given in the
CC specification (or complements/ fragments). The probe hybridises at high
CC stringency to a nucleic acid molecule expressed in the human adult liver.
CC (I) may be used for predicting, measuring and displaying gene expression
CC in samples derived from human adult liver. The genes identified may be
CC involved in genetic liver diseases such as cirrhosis,
CC hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which is
CC associated with coronary heart disease. ABG47348-ABG59930 represent human
CC liver single exon encoded peptides of the invention. Note: the sequence
CC information for this patent does not appear in the printed specification
CC but was obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 194 AA;

Query Match 10.1%; Score 1021; DB 4; Length 194;
Best Local Similarity 100.0%; Pred. No. 5.3e-59;
Matches 194; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1232 MFETSGDSLNCSTNDHCEQKEDLKEKNTNLFLOKPGSFSLKLELVAKMPPSEVMT 1291
DB 1 MFETSGDSLNCSTNDHCEQKEDLKEKNTNLFLOKPGSFSLKLELVAKMPPSEVMT 60

QY 1292 KPNAGANGCTLSYQNSGKHSIGSVQSTATQSNVEKADSNLNFNTGSSGPGKFYSPLPNDQ 1351
DB 61 KPNAGANGCTLSYQNSGKHSIGSVQSTATQSNVEKADSNLNFNTGSSGPGKFYSPLPNDQ 120

QY 1352 LKLTTEKRNQWFSLLRTPCDDTSLTHADNSTASLVTPQSQPPSKSPSPAPLGSSAQ 1411
DB 121 LKLTTEKRNQWFSLLRTPCDDTSLTHADNSTASLVTPQSQPPSKSPSPAPLGSSAQ 180

QY 1412 NPVGLNPFALSPLO 1425
DB 181 NPVGLNPFALSPLO 194

RESULT 24

AAM01830
ID AAM01830 standard; protein; 194 AA.
XX
AC AAM01830;
XX
DT 09-OCT-2001 (first entry)
XX
XX Peptide #512 encoded by probe for measuring human breast gene expression.
DE Probe; human; breast disease; breast cancer; development disorder;
KW inflammatory disease; proliferative breast disease; non-carcinoma tumour.
XX
XX Homo sapiens.
XX
XX WO200157270-A2.
XX
XX 09-AUG-2001.
XX
XX 29-JAN-2001; 2001WO-US000661.
XX
XX 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-476286/51.
XX
XX Novel single exon nucleic acid probe used to measuring gene expression in
PT a human breast.
XX
XX Claim 27; SEQ ID NO 10570; 322pp; English.
XX
XX The present invention relates to novel single exon nucleic acid probes
CC (see AAI00010-AA110067). The present sequence is a peptide encoded by one
CC such probe. The probes are useful for measuring human gene expression in
CC a human breast sample, where the probe hybridises at high stringency to a
CC nucleic acid expressed in the human breast. The probes are useful for
CC predicting, diagnosing, grading, staging, monitoring and prognosing
CC diseases of the human breast, particularly those diseases with polygenic
CC aetiology. The diseases include: breast cancer, disorders of development,
CC inflammatory diseases of the breast, fibrocystic changes, proliferative
CC breast disease and non-carcinoma tumours. Note: The sequence data for
CC this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 194 AA;

Query Match 10.1%; Score 1021; DB 4; Length 194;
Best Local Similarity 100.0%; Pred. No. 5.3e-59;
Matches 194; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1232 MFETSGDSLNCSTNDHCEQKEDLKEKNTNLFLOKPGSFSLKLELVAKMPPSEVMT 1291
DB 1 MFETSGDSLNCSTNDHCEQKEDLKEKNTNLFLOKPGSFSLKLELVAKMPPSEVMT 60

QY 1292 KPNAGANGCTLSYQNSGKHSIGSVQSTATQSNVEKADSNLNFNTGSSGPGKFYSPLPNDQ 1351
DB 61 KPNAGANGCTLSYQNSGKHSIGSVQSTATQSNVEKADSNLNFNTGSSGPGKFYSPLPNDQ 120

QY 1352 LKLTTEKRNQWFSLLRTPCDDTSLTHADNSTASLVTPQSQPPSKSPSPAPLGSSAQ 1411
DB 121 LKLTTEKRNQWFSLLRTPCDDTSLTHADNSTASLVTPQSQPPSKSPSPAPLGSSAQ 180

QY 1412 NPVGLNPFALSPLO 1425
DB 181 NPVGLNPFALSPLO 194

Db 181 NPVGLNPFALSPLO 194

RESULT 25
ABG35872
ID ABG35872 standard; peptide; 194 AA.
XX AC ABG35872;
XX DT 19-AUG-2002 (first entry)
XX DE Human peptide encoded by genome-derived single exon probe SEQ ID 25537.
XX KW Human; single exon probe; asthma; lung cancer; COPD; ILD;
KW chronic obstructive pulmonary disease; interstitial lung disease;
KW familial idiopathic pulmonary fibrosis; neurofibromatosis;
KW tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;
KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemorrhoidosis;
KW pulmonary histiocytosis; lymphangioleiomyomatosis; Karagener syndrome;
KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
KW primary ciliary dyskinesia; pulmonary hypertension;
KW hyaline membrane disease.
XX OS Homo sapiens.
XX PN WO200186003-A2.
XX PD 15-NOV-2001.
XX PF 30-JAN-2001; 2001WO-US000665.
XX PR 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DB;
XX WPI; 2002-114183/15.
XX DR Spatially-addressable set of single exon nucleic acid probes, used to
XX measure gene expression in human lung samples.
XX PS Claim 27; SEQ ID NO 25537; 634pp; English.
XX CC The invention relates to a spatially-addressable set of single exon
CC nucleic acid probes for measuring gene expression in a sample derived
CC from human lung comprising single exon nucleic acid probes having one of
CC 12614 nucleic acid sequences mentioned in the specification, or their
CC complements or the 12387 open reading frames derived from the 12614
CC probes. Also included are a microarray comprising the novel set of probes
CC; the novel set of probes which hybridise at high stringency to a nucleic
CC acid expressed in the human lung; measuring gene expression in a sample
CC derived from human lung, comprising (a) contacting the array with a
CC collection of detectably labeled nucleic acids derived from human lung
CC mRNA, and (b) measuring the label detectably bound to each probe of the
CC array; identifying exons in a eukaryotic genome, comprising (a)
CC algorithmically predicting at least one exon from genomic sequences of
CC the eukaryote; and (b) detecting specific hybridisation of detectably
CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,
CC having a fragment identical to the predicted exon, the probe is included
CC in the above mentioned microarray; assigning exons to a single gene,
CC comprising (a) identifying exons from genomic sequence by the method
CC above and (b) measuring the expression of each of the exons in several
CC tissues and/or cell types using hybridisation to a single exon
CC microarrays having a probe with the exon, where a common pattern of
CC expression of the exons in the tissues and/or cell types indicates that
CC the exons should be assigned to a single gene; a peptide comprising one

CC of 12011 sequences, mentioned in the specification, or encoded by the
CC probes/open reading frames (ORF). The probes are used for gene expression
CC analysis, and for identifying exons in a gene, particularly using human
CC lung derived mRNA and for the study of lung diseases such as asthma, lung
CC cancer, chronic obstructive pulmonary disease (COPD), interstitial lung
CC disease (ILD), familial idiopathic pulmonary fibrosis, neurofibromatosis,
CC tuberous sclerosis, Gaucher's disease, Niemann-Pick disease, Hermansky-
CC Pudlak syndrome, sarcoidosis, pulmonary haemorrhoidosis, pulmonary
CC histiocytosis, lymphangioleiomyomatosis, pulmonary alveolar proteinosis,
CC Karagener syndrome, fibrocystic pulmonary dysplasia, primary ciliary
CC dyskinesia, pulmonary hypertension and hyaline membrane disease. The
CC present sequence is a peptide/protein encoded by a single exon probe of
CC the invention. Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 194 AA;
Query Match 10.1%; Score 1021; DB 5; Length 194;
Best Local Similarity 100.0%; Pred. No. 5.3e-59;
Matches 194; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1232 MFETSGDSLNCSTNDHCEQKEDLKEKONTNLFLOKPGSFSLKLEVAKMPPESEVMT 1291
Db 1 MFETSGDSLNCSTNDHCEQKEDLKEKONTNLFLOKPGSFSLKLEVAKMPPESEVMT 60
QY 1292 KPNAGANGCTLSYQNSGKHSLGVSQSTATQSNVEKADSNLNFNTGSSGPGKFYSPLPNDQ 1351
Db 61 KPNAGANGCTLSYQNSGKHSLGVSQSTATQSNVEKADSNLNFNTGSSGPGKFYSPLPNDQ 120
QY 1352 LLKTLTEKNROWFSLPRTPCDDTSLTHADNSTASLVTTPSQSPSPSPPTAPLGSSAQ 1411
Db 121 LLKTLTEKNROWFSLPRTPCDDTSLTHADNSTASLVTTPSQSPSPSPPTAPLGSSAQ 180
QY 1412 NPVGLNPFALSPLO 1425
Db 181 NPVGLNPFALSPLO 194

RESULT 26
AAU16212
ID AAU16212 standard; protein; 160 AA.
XX AC AAU16212;
XX DT 07-NOV-2001 (first entry)
XX DE Human novel secreted protein, Seq ID 1165.
XX KW Human; immunosuppressive; antiarthritic; antirheumatic; cytostatic;
KW cardiant; vasotropic; cerebroprotective; neurotropic; neuroprotective;
KW antibacterial; virucide; fungicide; opthalmological; vulnery;
KW secreted protein; rheumatoid arthritis; hyperproliferative disorder;
KW cardiovascular disorder; cardiac arrest; cerebrovascular disorder;
KW cerebral ischaemia; angioneurosis; nervous system disorder;
KW Alzheimer's disease; infection; ocular disorder; corneal infection;
KW wound healing; epithelial cell proliferation; skin ageing; food additive;
KW preservative; antiproliferative.
XX OS Homo sapiens.
XX PN WO200155322-A2.
XX PD 02-AUG-2001.
XX PF 17-JAN-2001; 2001WO-US001341.
XX PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.

PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215133P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249219P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 06-DEC-2000; 2000US-0251719P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251989P.
PR 11-DEC-2000; 2000US-0251990P.
PR 05-JAN-2001; 2001US-0259678P.
XX
FA (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Baraah SC, Ruben SM;
XX
XX WPI; 2001-488783/53.
DR N-PSDB; AAS26199.
XX
PT New nucleic acid molecules encoding 461 human secreted proteins for
PT diagnosing, preventing, treating or ameliorating medical conditions and
PT used as food additives or preservatives.
XX
PS Claim 11; SEQ ID NO 1165; 980pp; English.
XX
XX The invention relates to isolated nucleic acid molecules and their
CC encoded secreted proteins. The nucleic acids and proteins are used to

CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,
CC rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used
CC in diagnosing a pathological condition or susceptibility to a
CC pathologic condition. Antibodies to the proteins can also be used in
CC alleviating symptoms associated with the disorders and in diagnostic
CC immunoassays e.g. radioimmunoassays or enzyme linked immunosorbant assays
CC (ELISA). Disorders which are diagnosed or treated include autoimmune
CC diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g.
CC neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac
CC arrest, cerebrovascular disorders e.g. cerebral ischaemia, angiogenesis,
CC nervous system disorders e.g. Alzheimer's disease, infections caused by
CC bacteria, viruses and fungi and ocular disorders e.g. corneal infection,
CC and many other disorders listed in the specification. The polypeptides
CC can also be used to aid wound healing and epithelial cell proliferation,
CC to prevent skin aging due to sunburn, to maintain organs before
CC transplantation, for supporting cell culture of primary tissues, to
CC regenerate tissues and in chemotaxis. The polypeptides can also be used
CC as a food additive or preservative to increase or decrease storage
CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,
CC minerals, cofactors and other nutritional components. The present
CC sequence represents a novel secreted protein of the invention. Note: The
CC sequence data for this patent did not form part of the printed

Query Match 8.1%; Score 817; DB 4; Length 160;
Best Local Similarity 98.7%; Pred. No. 1.1e-45;
Matches 156; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
QY 1815 TGDTEDEASSTSSIKRGNKDLQKQKBEENTSINLSKQESTSVKKPKRDSKDLALCS 1874
|||||
DB 3 TGDTEDEASSTSSIKRGNKDLQKQKBEENTSINLSKQESTSVKKPKRDSKDLALCS 62
QY 1875 MILTEMETHEDAWPFLLPVNLKLVPGYKVKVKKPMDFSTIREKSSGGYPNLETFLDVR 1934
|||||
DB 63 MILTEMETHEDAWPFLLPVNLKLVPGYKVKVKKPMDFSTIREKSSGGYPNLETFLDVR 122
QY 1935 LVFDCNCFNEDSDIGRAGHNMRYKFEKKWTDTFKVS 1972
|||||
DB 123 LVFDCNCFNEDSDIGRAGHNMRYKFEKKWTDTFKVS 160

RESULT 27
ABUS5281
ID ABUS5281 standard; protein; 160 AA.
XX ABUS5281;
XX AC
XX XX
DT 18-MAR-2003 (first entry)
XX
DE Human novel polypeptide #368.
XX
KW Human; neural disorder; immune system disorder; renal disorder;
KW muscular disorder; respiratory disease; reproductive disorder;
KW gastrointestinal disorder; pulmonary disorder; cardiovascular disorder;
KW hyperproliferative disorder; inflammatory disease; allergic reaction;
KW blood related disorder; cancer; immunosuppressive; antiinflammatory;
KW cardiovascular; nephrotropic; cytostatic; antiallergic; thrombolytic;
KW haemostatic; antiarteriosclerotic.
XX
OS Homo sapiens.
XX
PN US2002132753-A1.
XX
XX 19-SEP-2002.
XX
PF 17-JAN-2001; 2001US-00764864.
XX
XX 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 28-JUN-2000; 2000US-0214886P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.

PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 22-AUG-2000; 2000US-0226868P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 08-SEP-2000; 2000US-0231413P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 27-SEP-2000; 2000US-0235834P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241809P.
PR 01-NOV-2000; 2000US-0244617P.
PR 17-NOV-2000; 2000US-0249299P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
XX
XX (ROSE/) ROSEN C A.
PA (RUBE/) RUBEN S M.
PA (BARA/) BARASH S C.
XX
PI Rosen CA, Ruben SM, Barash SC;
XX
XX WPI; 2003-147444/14.
DR N-PSDB; ABX73540.
XX
PT New polypeptides and nucleic acids, useful in gene therapy for treating,
PT inhibiting or preventing e.g. neural, immune system, muscular,
PT respiratory, reproductive, gastrointestinal, pulmonary, cardiovascular or
PT renal disorders.
XX
XX Claim 11; SEQ ID NO 1165; 402pp; English.
PS
XX
XX The invention relates to human novel polypeptides and their associated
CC polynucleotides. The polypeptides and polynucleotides are useful in gene
CC therapy for treating, inhibiting or preventing neural disorders, immune
CC system disorders (e.g. systemic lupus erythematosus, rheumatoid arthritis
CC and multiple sclerosis), muscular disorders, respiratory diseases (e.g.
CC nasal vestibulitis, nasal polyps and sinusitis), reproductive disorders,
CC gastrointestinal disorders, pulmonary disorders, cardiovascular disorders
CC (e.g. congenital heart defects, Ebstein's anomaly and hypoplastic left
CC heart syndrome), renal disorders (e.g. acute kidney failure and end-stage
CC renal disease), hyperproliferative disorders (e.g. Hodgkin's disease and
CC leukaemia), inflammatory diseases (e.g. septic shock, bursitis and
CC appendicitis), allergic reactions and conditions (e.g. asthma), blood
CC related disorders (e.g. thrombosis, atherosclerosis and myocardial

CC infarction) and cancerous diseases. Sequences ABUS4914-ABUS5699 and
CC ABUS5748 represent human novel polypeptides of the invention
XX
SQ Sequence 160 AA;

Query Match 8.1%; Score 817; DB 6; Length 160;
Best Local Similarity 98.7%; Pred. No. 1.1e-45;
Matches 156; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1815 TGDTEDEDSASTSSSLKRGKDLQVKKMEENTSNLSKQESFTSVKKPKRDSKDLALCS 1874
DB 3 TGDTEDEDSASTSNLSKRGKDLKVKMEENTSNLSKQESFTSVKKPKRDSKDLALCS 62

QY 1875 MILTEMETHDAWPELLVNLKLVPGYKVKIKKMDFTIREKLSGGYPNLETFALDVR 1934
DB 63 MILTEMETHDAWPELLVNLKLVPGYKVKIKKMDFTIREKLSGGYPNLETFALDVR 122

QY 1935 LVFNCETFNEDDSIGRAGHNMRYFEKKWTDTFKVS 1972
DB 123 LVFNCETFNEDDSIGRAGHNMRYFEKKWTDTFKVS 160

RESULT 28
ID AAB21050 standard; protein; 555 AA.
XX AAB21050;
AC AAB21050;
XX
DT 19-DEC-2000 (first entry)
DE Human nucleic acid-binding protein, NuABP-54.
XX
KW Human nucleic acid-binding protein; NuABP; agonist; antagonist; EST;
KW expressed sequence tag; drug screening; recombinant expression; antibody;
KW reproductive disorder; infertility; immunological disorder;
KW neurological disorder; cell proliferative disorder; cancer; tumour.
XX
OS Homo sapiens.
XX
FN WO200044900-A2.
XX
PD 03-AUG-2000.
XX
PF 28-JAN-2000; 2000WO-US002237.
XX
PR 29-JAN-1999; 98US-0117904P.
PR 29-JAN-1999; 98US-0117905P.
XX
PA (INCY-) INCYTE PHARM INC.
XX
XX Tang YT, Lal P, Hillman JL, Yue H, Azimzai Y, Lu AMD, Baughn MR;
PI Tran B, Shih LL, Au-Young JL;
XX
XX WPI; 2000-499332/44.
DR N-PSDB; AAA72435.
XX
XX Novel nucleic acid binding proteins, used to identify agonists and
PT antagonists of them, for the treatment of reproductive, immunological,
PT neurological and cell proliferative disorders including cancer.
XX
PS Claim 1; Page 143-144; 180pp; English.
XX
XX Sequences AAB20997-B21051 represent novel human nucleic acid-binding
CC proteins (NuABPs) which are encoded by the cDNA sequences AAA72382-
CC A72436. The cDNAs were produced by extension from an appropriate EST
CC (expressed sequence tag) using primers designed using the EST. The
CC invention also relates to expression constructs, host cells and
CC transgenic organisms comprising a human NuABP nucleic acid, recombinant
CC production of the human NuABPs, and antibodies against the human NuABPs,
CC and also to methods of screening modulators of human NuABP activity or
CC expression. The human NuABPs, and their agonists and antagonists are used
CC to treat diseases associated with overexpression or underexpression of
CC functional NuABPs. Human NuABP proteins and nucleotides, and NuABP

CC agonists and antagonists can be used to diagnose, treat and prevent
CC reproductive, immunological, neurological and cell proliferative
CC disorders. Reproductive disorders that may be treated using compositions
CC of the invention include infertility, endometriosis, disruptions of the
CC menstrual cycle and disruptions of spermatogenesis. Immunological
CC disorders that may be treated include AIDS, allergies, and autoimmune
CC disorders such as multiple sclerosis, rheumatoid arthritis, diabetes and
CC systemic lupus erythematosus. Neurological disorders that may be treated
CC include epilepsy, neurodegenerative conditions such as Alzheimer's
CC disease and Parkinson's disease, prion diseases such as Creutzfeldt-Jakob
CC disease, and mental disorders such as schizophrenia. Cell proliferative
CC disorders that may be treated include a wide variety of cancers, and also
CC arteriosclerosis, atherosclerosis, cirrhosis and psoriasis
XX
SQ Sequence 555 AA;

Query Match 7.5%; Score 762.5; DB 3; Length 555;
Best Local Similarity 31.5%; Pred. No. 3e-41;
Matches 185; Conservative 104; Mismatches 173; Indels 125; Gaps 12;

QY 388 AKSLVEQFRGTDSIPSKSDSESDDEDEDEDEDEDEDEDEDEDEDEDEDEDEDESDT 447
DB 28 AFSLLADDSQTSIFASPTSPVVGESVLQDNSFDLNGSDAQEEMETOSSDFFPPLST 87

QY 448 EGSEEDDDDDKQDSDSDEGEKTSMLKNTTS----- 481
DB 88 QPÄ-----PQSSITQLHPATSPAVSTTSPAVSLVSPASPEISPEVCPAASTV 138

QY 482 -----SVKGPSMS-----LTGHSPTPRNLHIAPGSA-PAALCESOSAPPLG 523
DB 139 VSPAVFVSVPASSAVLPAVSLEVLTAASVTPSPASPVTPSPAAAPTASPANKDVSSPLE 198

QY 524 TSSST-----LTSSPSHGTSKRRVTDRELAIPLEYGHQRETRTNFGRLOGEVAY 576
DB 199 TTADVBEITGELTAS-GSGDWMRRRTATPEEVLPLQHGWRVRVIRKNSHRWQGETWY 257

QY 577 YAPCGKLRQYPEVIKYLRSNGIMDISDNFSAKIRVGDYFEBARDGPQWQWCLLKEE 636
DB 258 YGPGCKRMKQPEVIKYLRSNVHSVREHFSFPRMPVGVDFPERDTPGELQWVLSAE 317

QY 637 DVIPRIAMEGRGRPNPDQRAREESRMRMRKGRPNVGNABFLDNADAKLRKLOAQ 696
DB 318 EIPSRIOAITGKGRPRNTEKATKEVPKVRGRGRPPKVTITELLNTDNRPLKLEAQ 377

QY 697 BIARQAQIKLLRLKQEQARVAKEAKQ-----QAIMAAEKKKQEQIKIMKQ 747
DB 378 E-----TLNEEDKAKIAKSKKQKQVQGEQCTTIQQAARKKRQETKSLKQ 425

QY 748 QEKIKRIQIRMEKELRAQQLLEAKKKKEAANAALLLEAEKRIKEKEMRRQQAALLKHQ 807
DB 426 K-----EAKKSKAEKKGK--TKQEKLEK-----VKRE 453

QY 808 ERERRRQHMLMKAMEARKKAEKKERLKQEKRDEKRLNKKERKLBORRLELEMAKELKPN 867
DB 454 KKEKVQM-----KEKEEVTAKPACKADKTLATQRLBERQORQOMILEDMKPT 502

QY 868 EDMCLADOKPLPELPRIPLGLVLSGTSFSDCLMVVQFLNFGKVLGFD 914
DB 503 EDMCLTDHQLPDRFVPLGLTLPSCAFSDCLTIVEFLHSFGKVLGLD 549

RESULT 29
AAAY07734
ID AAAY07734 standard; protein; 1540 AA.
XX
XX AC AAAY07734;
XX
XX DT 02-JUL-1999 (first entry)
XX
XX DE Human hTLP2 protein.
XX
XX KW Telomerase protein; hTLP2; human; biocontrolling mechanism; cell growth;
XX cell aging.

XX OS Homo sapiens.
XX PN JP11089579-A.
XX PD 06-APR-1999.
XX PF 24-SEP-1997; 97JP-00258450.
XX PR 24-SEP-1997; 97JP-00258450.
XX PA (MITU) MITSUBISHI CHEM CORP.
XX WPI: 1999-290828/25.
XX DR N-PSDB; AAX37304.
XX PT A higher animal telomerase protein and a gene coding it.
XX PS Example 1; Page 12-18; 22pp; Japanese.
XX CC This invention describes a novel human telomerase protein, hTLP2. This
XX CC protein and the gene coding it are useful for the elucidation of
XX CC biocontrolling mechanisms such as cell growth and cell aging
XX SQ Sequence 1540 AA;

Query Match 7.2%; Score 727.5; DB 2; Length 1540;
Best Local Similarity 21.3%; Pred. No. 2.9e-38;
Matches 340; Conservative 182; Mismatches 424; Indels 647; Gaps 53;

QY 664 SMRRKGRPPVGNVNAEFLDNADAKLLRLQAEIARQAQIILRLKQKQOARVAKEA 723
DB 270 SPANRRGRPP-----KRIHISQEDNVA--- 292

QY 724 KQQAIAAEKREKKEQIKIMKQOEKIKIOIRMEKELRAQILEAKKKKKEAANAK 783
DB 293 --NKOTLASYSKATKEDKILKQEE---MKSLAF-----KAKLRKAKDA- 334

QY 784 LLEAKRIKEKEMRRQQAALLKHQERRRRQHMLMKAMEARKKAEKERLKOERDEKR 843
DB 335 -LEAKK--KEKE-----DKEKKREELKXIVEERLKKKEERLAVEREKE-- 377

QY 844 LNKERKLQRRLLEMAKELKPNEDMCLADOKPLPELPRIGLVLSGSTFSDCLMVQF 903
DB 378 --REKLREKRYKIVYLQWSPREDMECDLKEPPE-PTPVKTRLPEIFGDALMVLEF 434

QY 904 LNFQKVLGFDVNIQVN--LSVLQEGILNIGDSMEVODLLVRLLSAAV----- 951
DB 435 LNAFQEL--FDLQDFPDQVTLVELEALVG--NDSEGLPCELLFFFLTAIFAQIAEBE 491

QY 952 -----COPGLITGYKAKTAL-----GEHLNMGV 975
DB 492 VAKEQLTDADTKDLTEALDEDADP-----TKSALSASVASLAAPQLHQCSLSDL 544

QY 976 NRDNVSEILQIF-----MEAHCGGTETTESLTKYA 1005
DB 545 DCSLTSELRLHLASGADVTSANAKRYQKRGFGFDATDDACMELRLSNPSVLKLSSTS 604

QY 1006 FOAHTPAOKASVLAFLINELACSKSVSEIDKNIDYMSNLRDKWV----- 1052
DB 605 VYDLTPGEMKILHALCGKLL---TLVSTRDFIEDYVILRLQAKQEFRELKABQHRKERE 661

QY 1053 -----EGKLR-----KLRIHAKTKGTRDTSGGIDIGEBQHPGLTPTGKRERR 1096
DB 662 EAAARIKREKEKLEKEQKQKKEKQELKEDQNRNSTADISIGEEE-----R 708

QY 1097 KGGDSYDDDDDDDDGDEDEDEED-----KEDQKQK-----KT 1133
DB 709 EDFDTSIESKQTEQKELQDMVTEDEDDPGSHKRRGRKRGKQNGFKETROEQINCVTRE 768

QY 1134 DICEDEGDQAAASVEELEKQLEKLSKQSQVRRKLPDASHSLRSVMFGPDYRRRYWIL 1193
DB 769 PLTADEEE---ALKQEHORKEKELLEKIQS-----AIACNTIFPLGRDMRYRYWIF 817

QY 1194 PRCGIFVEGMESGEGLEIEAKERBKLKKAESVQIKKEMFETSGDSLNCNTDHCQKED 1253
DB 818 PSIPGLFIE--EDYSGLTE----- 834

QY 1254 LKEKONTNLFLOKPGSFSKLSKLLLEVAKMPPESEVMTPKPNAGANGCTILSYQNSGKHSLG 1313
DB 835 -----DMLLPRPSSFQ-----NVQSDPQVSTKTGEP----- 862

QY 1314 SVQSTATQSVNEKADSNLNFNTGSGPGKFKYFPLND---QLLKTLTEKNRWFSLPRT 1370
DB 863 -LMSEST-SNIDQG-----PRDHSVQLPKPVHKENR-W----- 892

QY 1371 PCDDTSLTHAMSTASLTVPQSPSKSPSPPTAPPLGSSAQNVPGLNPPALSPLOVKGGV 1430
DB 893 -CFYSSCEQLDLIEAL----- 908

QY 1431 SMWGLQFCGWPTGVVTSNIPFTLSVPSLGSGLSEGNCSFLTNSNVASSKSESPPQNE 1490
DB 909 -----NSRGHRESALKETLLQEKSR----- 928

QY 1491 KATSAQPAAVEVAK--PVDFFSPKPIPEMOPFGWRIIDPEDLKALLKVLHLR-----G 1542
DB 929 --ICAQLARFSEKEKPFSDKQPDSPKPTYSRGRSSNAYDPSQMC-A-EKQLELRDLRDL 985

QY 1543 IREKALQ-----KOIQKHLDYITQACLKNKDVAIIELNENENENOVTRDIVENWSVEQA 1596
DB 986 IEDRIYQGTGAIKVTDHR--IWSALESGRYELLS--EENKENGIIKTV--NEDVEEME 1039

QY 1597 MEMDLSVLQOVEDLERRVVASASLVQGMWCPEPASEREDLVYFEHKSFTKLCKEHDGET 1656
DB 1040 IDEQTKVIVKORLLGIKTETPTSTVNASTPQSVS----- 1074

QY 1657 GEDESSAHALERKSDNPLDIAVTRLADLERNIETRIEEDAP-----GLRVVR 1704
DB 1075 -----SVVHYL-----AMALFOIEQGIERRFLK--APLDASDSGRSRYTVLDWR 1117

QY 1705 RALSARSAAQVALCICQOLQKSIWEKSIKMYCOICRKGDNELLLLLCDGCKGCHTYC 1764
DB 1118 ESSLSSASLSQVFLHLSTLDRSVIWSKILNARCKICRKGDAENWVLCDCGDRGHHTYC 1177

QY 1765 HRPKTTITPDGWFPCACIAGSGOTLAKLH--VKGKKTNESKKGKVTLTGTE--- 1819
DB 1178 VRPKLKTVPBGDFWFCPCRPQSRRLSRQPSLEDEDEDSMGDEDEVDGDEEGQ 1237

QY 1820 -----DED-----SASTSS 1828
DB 1238 SEEBEYEQDEDDSRPLPVKTRGKLSSFSRSGQQQEPGRYPSRQQSTPTKTTVSKTGR 1297

QY 1829 SL-----KRGNKD----- 1836
DB 1298 SLRKINSAPPTETKSLRIASRSTRSHGHPLOADVVELLSPPRRKRGRKSANNTTPENS 1357

QY 1837 -----LOKRKVEENTSIN---LSQBSFTSVKPKRQDSK----- 1868
DB 1358 FPNFRVIATKSESQSRSVNIASKLQSESKRCKRQSPSPVTLGRSSRGGGVH 1417

QY 1869 DLALCSMLTMEWETHEDAWPFLPVNLKLVPCYKVIKKPMDPSTIREKLSGGQVPLNET 1928
DB 1418 ELSAFEQVULVELRHDDSWPFLKLVSKIQVDDYDIIKKPIALNIIREKNVCKEYKLASE 1477

QY 1929 FALDVLVDFDCETFNEDDDSDIGRAGHNMRYF 1961
DB 1478 FIDDIELMFSNCFEYNPNTSEAKAGTRLOAFF 1510

RESULT 30

AAW81169

ID AAW81169 standard; protein; 1674 AA.

XX AC AAW81169;

XX DT 05-MAY-1999 (first entry)


```

QY 1887 WPFLPVNLKLVPGYKVKIKPMDESTIREKLSGQYNLETFALDVRLVDFNCSTFNE 1946
DB 1570 WPFLKLVSKIQVDPDYDIKPKPIALNIRKVNKCEYKLASEFIDIELMFSCNCFEYNPR 1629
QY 1947 DSDIGRAGHMRKYF 1961
DB 1630 NTSEAKAGTRLQAF 1644

RESULT 31
ID ADF69139
XX ADF69139 standard; protein; 1674 AA.
AC ADF69139;
XX 12-FEB-2004 (first entry)
XX Human MP53 protein sequence SEQ ID NO:109.
XX p53 pathway modulating agent; MP53; p53 modulator; cytostatic;
KW gene therapy; cancer; human.
XX Homo sapiens.
OS
XX WO2003083047-A2.
XX 09-OCT-2003.
XX 28-FEB-2003; 2003WO-US006025.
XX 01-MAR-2002; 2002US-0361196P.
XX (EXEL-) EXELIXIS INC.
XX Belvin M, Francis-Lang H, Friedman L, Plowman GD, Heuer TS, Li D;
PI Funke RP;
XX WPI; 2003-812540/76.
DR N-PSDB; ADF69195.
XX
PT Identifying a candidate p53 pathway modulating agent for treating e.g.,
PT cancer by contacting an assay system comprising a MP53 polypeptide or
PT nucleic acid with a test agent and detecting a test agent-biased
PT activity.
XX
PS Example; SEQ ID NO 109; 406pp; English.
XX
CC The present invention describes a method for identifying a candidate p53
CC pathway modulating agent, which comprises: (a) providing an assay system
CC comprising a MP53 (modulator of p53) polypeptide or nucleic acid or its
CC fragment or derivative; (b) contacting the assay system with a test agent
CC under conditions where the system provides a reference activity except in
CC the presence of the test agent; and (c) detecting a test agent-biased
CC activity, where a difference between the test agent-biased activity and
CC the reference activity identifies the test agent as a candidate p53
CC pathway modulating agent. Also described: (1) modulating the p53 pathway
CC of a cell; (2) modulating the p53 pathway in a mammalian cell; and (3)
CC diagnosing a disease in a patient. MP53 has cytostatic activity, and can
CC be used in gene therapy. The method is useful for identifying a candidate
CC p53 pathway modulating agent for preparing a composition for diagnosing
CC or treating e.g., cancer. The present sequence represents a human MP53
CC protein, which is used in the exemplification of the present invention.
XX
SQ Sequence 1674 AA;
Query Match 7.2%; Score 726.5; DB 7; Length 1674;
Best Local Similarity 21.5%; Pred. No. 3.8e-38;
Matches 338; Conservative 186; Mismatches 424; Indels 627; Gaps 53;
QY 664 SRMRRRGRPNVGNAEFLDNADAKLRLKLAQETARQAQIKLRLKQKQEQARVAKEA 723
DB 420 SPANRRGRPP-----KRIHSQEDNVA--- 442

```

```

QY 724 KQQAIIAAEKRKQKQEQIKIMKQEQIKRIQOIRMEKELRAQOILEAKKKKEEAAAK 783
DB 443 --NQOTLASYSRKATKDKLQKE-----MKSLAFE-----KAKLKREKADA- 484
QY 784 LLEAEKRIKEKEMRRQOAVLLKHQERERRRRHMLMKAMAEARKKAAEERLKKQEKDEKR 843
DB 485 -LEAKK--KEKK-----DKKKREBELKIVEEERLKKKEERLKKVEREK-5-- 527
QY 844 LNKERKLEORLELEMAKELKKPNEDMCLADQKLPPLPGLVLSGSTFSDCLMVVQF 903
DB 528 --REKLREKRYEYLKQWSKPREDMCEDDLKELPE-PTPVKTRLPPEIFGDALMVLEF 584
QY 904 LRNFQKVLGFDVNIIDVPN--LSVLQEGLLNIGDSMGEVDLLVRLLSA---AVCDPLGI 957
DB 585 LNAFQEL--FDLQDEFDPDVTLEVEALVG-NDSGPLCELLFPFLTAIFAIAEEREE 641
QY 958 TGYKAKT---ALGEHLLNVGNVRDNVSEILQIF----- 987
DB 642 VAKQLTADATKGCSLKSLDLDSCTLSILRLHLILASGADVTSANAKRYQKRGFDATD 701
QY 988 --MEAHCGQTELTESLKTAKFAQHTPAQKASVLAFLINELACSKSVYSEIDKNIDYMSN 1044
DB 702 DACHELRLSNPSLVKKLSSTSVYDLTPGEKWKILHALCCKLL---TLVSTRDFIEDYDI 758
QY 1045 LRDRKVVY-----EGKLR-----KLRIHAKTKGRDTSGG 1075
DB 759 LRQAKQEPRELKAFQHRKEREAAAAIRKKEEKLKEQEQMKKEQKELKEDEQNSTAD 818
QY 1076 IDLGEBOHPLGTPPTGKRKRKGGSDYDDDDDDDDQDQDEDEDEDEDEDEDEDEDEDEDE 1131
DB 819 ISIGEEE-----REDFDTIESKDTQKELDQDMFTEDDDPGSHKRGREGK 865
QY 1132 KTD-----ICEDEDEGQAAASVELEKQIEKLSKQOQYRRKLPDA 1172
DB 866 RQNGFKFETROEQINCVTRELLTADEEE---ALKQEHQKEKELEKIQS-----A 914
QY 1173 SHLSRVNMFGRDYYRRYRWILPRCGGIFVEGMESEGLEETAKEREKLLKAAESVQIKEM 1232
DB 915 IACTNIPPLGRDRMYRWYIFPSIPGLFIE--EDYSGLTE----- 952
QY 1233 PETSGDSLNCNTDHCQEKEDLKEKDNVNLQKPGFSKLSKLELVAKMPESVMTPK 1292
DB 953 -----DMLLPPPSFQN-----NVQSDPQVSTKTGE 979
QY 1293 PNAGANGCTLSYQNSKHSLSGVOSTATQSNVEXADSNLNFTGSSGPGKFPYPLPND-- 1350
DB 980 P-----LMSEST-SNIDQ-----PRDHS 997
QY 1351 -QLLKLTLTEKNRWFSLLPRTPCDDTSLTHADMSTASLVTPSQSPSPSPPTAPLGS 1409
DB 998 VQLPKPVHKPNR-W-----CFYSSCEQLDLQLEAL----- 1026
QY 1410 AONPVGLNPFALSPLOKVGVMGLOFCGWPTGVVTSNIPFTLSVPSLSGLSGLSENG 1469
DB 1027 -----NSRGR 1032
QY 1470 NSPLTSNVASSKSESPPVQNEKATSAQPAAVEVAK--PVDFPSKPIPEEMQFGWRIID 1527
DB 1033 ESALKETLLQEKSR-----ICQLARFSEBKHFSDKPDQDSKFTYSGRSSNAYD 1083
QY 1528 PEDLKALLKVLHLR-----GIREKALQ-----KQTKHLDYITQACLKNKDVAIELN 1575
DB 1084 PSQWCA-EKQLELRDLDFLLDIEDRIYQGTGALGKVTDRH---IWRSALESGRVELLS-E 1138
QY 1576 ENEENQVTRDIVENWSVEEQAMEMDLVLOQVEDLERRVASASLOVGMWCPESAPERED 1635
DB 1139 ENKENGIIKTV--NEDVEEMEIDEQTKVIVKDLRGINKITETPTSTVSTNASTPQSVS---- 1192
QY 1636 LVYFEHKSFTKLCHEHGEFTGEDESSNAHALERKSDNPLDIAVTFLADLERNIERRIED 1695
DB 1193 -----SVVHYL-----AMALFOIEGIERRRFLK- 1215

```


QY 1410 AONPVLNPFALSPLOVKGVSMMGLQFCGWPTGVVTSNIPFTLSVPSLGSGLG:SEGNQ 1469
Db 1027 -----NSRGHR 1032
QY 1470 NSFLTSNVASSKESPPVQNEKATSAQPAAVEAK--PVDFFSPKPIPEEMQFGWRIID 1527
Db 1033 ESALKETLQEKSR-----ICQLARFSEKPHFSDKFPQDSKPTYSRGRSNAYD 1083
QY 1528 PEDLKALKLVHLR-----GIREKALQ-----KQIQKHLDYITQACIKNDKNDVAILN 1575
Db 1084 PQOMCA-EKQLELRURDFLLDIEDRIYQGTGLAIVTDPH-----IWRSALESRYELLS-E 1138
QY 1576 ENEENQVTRDIVENSVBEQAMEMDLVLUQVEDLERRVASLQVKGMMCPPEASERED 1635
Db 1139 ENKENGIIKTV--NEDVEEMEIDEQTKVIVKDRLLGIKTETPTSTVSTNASTPQSVS---- 1192
QY 1636 LVYFEHFKTKLCKEHDGEFTGEDSSAHLERKSDNPLDIAVTBLADLERNIERIBED 1695
Db 1193 -----SVHYL-----AMALFOIEQGIERRFLK- 1215
QY 1696 IAP-----GLRWRRALSEARSAAQVACIQOQKSIWAEXSIMKVQICIRK 1743
Db 1216 -APLDASDSGRSYKTVLDWRRESLSSASLSQVFLHLSTLDRSVIWSKILNARCICKR 1274
QY 1744 GNEELLLLDCGDCGCHTYCHRPKITTPDGDWFCPACIAKASGQTLKIKLHVKGKKT 1803
Db 1275 KGDAENMVLDCGDCRGHHTYCVRPKLKTVPEGDFCPECRPKQ-----RCRRLSPRQPS 1329
QY 1804 NESKK-----GKKVTLTGDTF-----DED----- 1822
Db 1330 LESDEDEDSMGDEDDVDGDEEQSBEEVEVEQDEDDSOEEBESVLPKRGPRQVRLP 1389
QY 1823 -----SASTSSL----- 1830
Db 1390 VKTRKLSLSSFSRQOQEPGRYPSRQOSTPKTVSSKTSRSLKINSAPPTKTSRI 1449
QY 1831 -----KRGKND-----LQKRMBSNTSI 1848
Db 1450 ASRSTRSHGPLQADVVELLSPRKRGRKSNANTPENSPPNFRVIAATKSSQSRSV 1509
QY 1849 N-----LSQESFTSVKKPKRDSK-----DLALCSMLTETHED 1886
Db 1510 NIASKLSLQESSEKRCRKQSPPEPVTGLRRSSGROGGVHELSAFEQLVVELVRHDS 1569
QY 1887 WFLPLVNLKLVPGYKVKIKPMDFSTIREKLSGOYPNLETFALDVLDPDNCETENED 1946
Db 1570 WFLKLVSQIQVPDYDIKKPIALNIIEKVNKCEYKLASEFIDDIELMFSNCFEYNPR 1629
QY 1947 DSDIGRAGHNMRYF 1961
Db 1630 NTSEAKAGTRLQAFF 1644

RESULT 33
ID ABU70931
XX ABU70931 standard; protein; 280 AA.

AC ABU70931;
XX

DT 10-JUN-2003 (first entry)
XX

DE Human adipocyte Selected Interacting domain, SID, #562.
XX
KW Human; prey; adipocyte; SID; selected interacting domain; anorectic;
KW antidiabetic; protein-protein interaction; diabetes;
KW yeast 2-hybrid assay; metabolic disorder; obesity.

XX Homo sapiens.
OS

PN WO200286122-A2.
XX

PD 31-OCT-2002.
XX

XX 14-MAR-2002; 2002WO-EP003768.
XX
XX 14-MAR-2001; 2001US-0275734P.
XX
XX (HYBR-) HYBRIGENICS.
XX
XX Legrain P, Daviet L;
XX
XX WPI; 2003-103412/09.
XX N-PSDB; ACA57475.
XX
XX New complex between two interacting proteins in adipocyte cells, useful
XX for identifying selected interacting domains that modulate protein
XX interactions, or for preventing or treating metabolic disorders such as
XX obesity or diabetes.
XX
XX Claim 6; Page 291-292; 382pp; English.

XX The invention relates to a complex between two interacting proteins in
XX adipocyte cells, given in the specification. The proteins are identified
XX by selecting a bait protein from a known adipocyte marker and then
XX performing a yeast 2-hybrid selection to isolate prey proteins encoded by
XX members of an adipocyte cDNA library. The proteins are designated as
XX (RTM) (selected interacting domains) proteins. Also included are a
XX polynucleotide encoding a polypeptide in the adipocyte cells, a
XX recombinant host cell expressing at least one of the interacting
XX polypeptides of the complex, selecting a modulating compound in adipocyte
XX cells, a SID (RTM) polypeptide comprising any of the 738 amino acid
XX sequences given in the specification (including its fragment or variant),
XX a SID (RTM) polynucleotide comprising any of the 738 nucleotide sequences
XX given in the specification (including its fragment or variant), a vector
XX comprising the SID (RTM) polynucleotide, a recombinant host cell
XX comprising the vector, a protein chip comprising the polypeptides and a
XX record comprising all or part of the data, listed in the specification.
XX The complex, polypeptides, polynucleotides and compounds are useful for
XX preventing or treating metabolic disorders such as obesity or diabetes.
XX The polynucleotides are useful as probes or primers. The complex is
XX particularly useful for identifying selected interacting domains (SID
XX (RTM)) for screening drugs that modulate the protein interaction, thus
XX exhibiting the therapeutic effect. The present sequence represents a SID
XX (prey) protein of the invention

XX Sequence 280 AA;

Query Match 6.3%; Score 638.5; DB 6; Length 280;
Best Local Similarity 45.6%; Pred. No. 1.6e-33;
Matches 129; Conservative 62; Mismatches 85; Indels 7; Gaps 1;

QY 793 EKEMRQOAVLLKHQERRRRQHMLMKAMEARKKAEERLKQEKRDKERINKERKLEQ 852
Db 4 EKEGKTKQEKLEKVKREKKEVKQM-----KEKEVTKAKPACKADKATLTATQRLEE 56
QY 853 RRLEEMAKELKKNEDMCLADQKPLPPLPGLVLSGTFSDCLMVVQFLNFGKVLG 912
Db 57 RQKQOMILEENKKXPTEDMCLTDHOPLPDFSRVPGLTLPAGAFSDCLTIVEFLHSFGKVLG 116
QY 913 FDNVDVNLVLOEGLNIGDSMGVEDLLVRLLSAAVCDPGLITGYKAKTALGEHLN 972
Db 117 FDPKDVPSLGVLOEGLLCQGDLSGEVDLLVRLKAAALHDPGFPFSCSLKILGEKVS 176
QY 973 VGVNRDNYSEILQIFMEAHCGQTELTESLTKTAKFOAHTPAQKASVLAFLINELACKSVV 1032
Db 177 IPLTRDNYSEILRCFLMAYGVEPALCDRLRTPQAPPPQKAAVLAFPVHELNGSTLII 236
QY 1033 SEIDKNIDYMNLRDRKRVGKURKURIIHAKTKTGKEDTSGG 1075
Db 237 NEIDKTLSEMSYRKNKWIIEGRLRLRLKTVLAKRTGRSEVMG 279

RESULT 34
ID AAU16625
AAU16625 standard; protein; 134 AA.

XX AAU16625; PR 08-SEP-2000; 2000US-0231141P.
XX AC 07-NOV-2001 (first entry) PR 08-SEP-2000; 2000US-0232080P.
XX DT 07-NOV-2001 (first entry) PR 08-SEP-2000; 2000US-0232081P.
XX DE 07-NOV-2001 (first entry) PR 12-SEP-2000; 2000US-0231968P.
XX KW Human novel secreted protein, Seq ID 1578. PR 14-SEP-2000; 2000US-0232397P.
XX KW Human; immunosuppressive; antiarthritic; antirheumatic; cytostatic; PR 14-SEP-2000; 2000US-0232399P.
KW cardiant; vasotropic; cerebroprotective; nootropic; neuroprotective; PR 14-SEP-2000; 2000US-0232400P.
KW anticardial; virucide; fungicide; ophthalmological; vulnerary; PR 14-SEP-2000; 2000US-0232401P.
KW secreted protein; rheumatoid arthritis; hyperproliferative disorder; PR 14-SEP-2000; 2000US-0233063P.
KW cardiovascular disorder; cardiac arrest; cerebrovascular disorder; PR 14-SEP-2000; 2000US-0233064P.
KW cerebral ischaemia; angiogenesis; nervous system disorder; PR 21-SEP-2000; 2000US-0234223P.
KW Alzheimer's disease; infection; ocular disorder; corneal infection; PR 21-SEP-2000; 2000US-0234274P.
KW wound healing; epithelial cell proliferation; skin ageing; food additive; PR 25-SEP-2000; 2000US-0234957P.
XX preservative; antiproliferative. PR 25-SEP-2000; 2000US-0234958P.
XX Homo sapiens. PR 26-SEP-2000; 2000US-0235484P.
XX WO200155322-A2. PR 27-SEP-2000; 2000US-0235834P.
XX PN 02-AUG-2001. PR 27-SEP-2000; 2000US-0235836P.
XX PD 02-AUG-2001. PR 29-SEP-2000; 2000US-0236327P.
XX PF 02-AUG-2001; 2001WO-US001341. PR 29-SEP-2000; 2000US-0236327P.
XX PF 31-JAN-2000; 2000US-0179065P. PR 29-SEP-2000; 2000US-0236367P.
PR 04-FEB-2000; 2000US-0180628P. PR 29-SEP-2000; 2000US-0236368P.
PR 24-FEB-2000; 2000US-0184664P. PR 29-SEP-2000; 2000US-0236369P.
PR 02-MAR-2000; 2000US-0186350P. PR 02-OCT-2000; 2000US-0236802P.
PR 16-MAR-2000; 2000US-0189874P. PR 02-OCT-2000; 2000US-0237037P.
PR 17-MAR-2000; 2000US-0190076P. PR 02-OCT-2000; 2000US-0237038P.
PR 18-APR-2000; 2000US-0198123P. PR 02-OCT-2000; 2000US-0237039P.
PR 19-MAY-2000; 2000US-0205515P. PR 13-OCT-2000; 2000US-0237040P.
PR 07-JUN-2000; 2000US-0209467P. PR 13-OCT-2000; 2000US-0239935P.
PR 28-JUN-2000; 2000US-0214886P. PR 20-OCT-2000; 2000US-0240960P.
PR 30-JUN-2000; 2000US-0215135P. PR 20-OCT-2000; 2000US-0241221P.
PR 07-JUL-2000; 2000US-0216647P. PR 20-OCT-2000; 2000US-0241785P.
PR 07-JUL-2000; 2000US-0216800P. PR 20-OCT-2000; 2000US-0241826P.
PR 11-JUL-2000; 2000US-0217487P. PR 01-NOV-2000; 2000US-0244617P.
PR 11-JUL-2000; 2000US-0217496P. PR 08-NOV-2000; 2000US-0244647P.
PR 14-JUL-2000; 2000US-0218290P. PR 08-NOV-2000; 2000US-0246475P.
PR 26-JUL-2000; 2000US-0220963P. PR 08-NOV-2000; 2000US-0246476P.
PR 26-JUL-2000; 2000US-0220964P. PR 08-NOV-2000; 2000US-0246477P.
PR 14-AUG-2000; 2000US-0224518P. PR 08-NOV-2000; 2000US-0246478P.
PR 14-AUG-2000; 2000US-0224519P. PR 08-NOV-2000; 2000US-0246523P.
PR 14-AUG-2000; 2000US-0225213P. PR 08-NOV-2000; 2000US-0246524P.
PR 14-AUG-2000; 2000US-0225214P. PR 08-NOV-2000; 2000US-0246525P.
PR 14-AUG-2000; 2000US-0225266P. PR 08-NOV-2000; 2000US-0246526P.
PR 14-AUG-2000; 2000US-0225267P. PR 08-NOV-2000; 2000US-0246527P.
PR 14-AUG-2000; 2000US-0225268P. PR 08-NOV-2000; 2000US-0246528P.
PR 14-AUG-2000; 2000US-0225270P. PR 08-NOV-2000; 2000US-0246532P.
PR 14-AUG-2000; 2000US-0225447P. PR 08-NOV-2000; 2000US-0246609P.
PR 14-AUG-2000; 2000US-0225757P. PR 08-NOV-2000; 2000US-0246610P.
PR 14-AUG-2000; 2000US-0225758P. PR 08-NOV-2000; 2000US-0246611P.
PR 14-AUG-2000; 2000US-0225759P. PR 08-NOV-2000; 2000US-0246613P.
PR 18-AUG-2000; 2000US-0226279P. PR 17-NOV-2000; 2000US-0249207P.
PR 22-AUG-2000; 2000US-0226681P. PR 17-NOV-2000; 2000US-0249208P.
PR 22-AUG-2000; 2000US-0226868P. PR 17-NOV-2000; 2000US-0249209P.
PR 22-AUG-2000; 2000US-0227182P. PR 17-NOV-2000; 2000US-0249210P.
PR 23-AUG-2000; 2000US-0227009P. PR 17-NOV-2000; 2000US-0249211P.
PR 30-AUG-2000; 2000US-0228924P. PR 17-NOV-2000; 2000US-0249212P.
PR 01-SEP-2000; 2000US-0229287P. PR 17-NOV-2000; 2000US-0249213P.
PR 01-SEP-2000; 2000US-0229343P. PR 17-NOV-2000; 2000US-0249214P.
PR 01-SEP-2000; 2000US-0229344P. PR 17-NOV-2000; 2000US-0249215P.
PR 01-SEP-2000; 2000US-0229345P. PR 17-NOV-2000; 2000US-0249216P.
PR 05-SEP-2000; 2000US-0229509P. PR 17-NOV-2000; 2000US-0249217P.
PR 05-SEP-2000; 2000US-0229513P. PR 17-NOV-2000; 2000US-0249218P.
PR 06-SEP-2000; 2000US-0230437P. PR 17-NOV-2000; 2000US-0249244P.
PR 06-SEP-2000; 2000US-0230438P. PR 17-NOV-2000; 2000US-0249245P.
PR 08-SEP-2000; 2000US-0231242P. PR 17-NOV-2000; 2000US-0249264P.
PR 08-SEP-2000; 2000US-0231243P. PR 17-NOV-2000; 2000US-0249265P.
PR 08-SEP-2000; 2000US-0231244P. PR 17-NOV-2000; 2000US-0249266P.
PR 08-SEP-2000; 2000US-0231413P. PR 17-NOV-2000; 2000US-0249297P.

PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-488783/53.
DR N-PSDB; AAS26612.
XX
XX New nucleic acid molecules encoding 461 human secreted proteins for
PT diagnosing, preventing, treating or ameliorating medical conditions and
PT used as food additives or preservatives.
XX
XX Claim 11; SEQ ID NO 1578; 980pp; English.
XX
CC The invention relates to isolated nucleic acid molecules and their
CC encoded secreted proteins. The nucleic acids and proteins are used to
CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,
CC rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used
CC in diagnosing a pathological condition or susceptibility to a
CC pathological condition. Antibodies to the proteins can also be used in
CC alleviating symptoms associated with the disorders and in diagnostic
CC immunoassays e.g. radioimmunoassays or enzyme linked immunosorbent assays
CC (ELISA). Disorders which are diagnosed or treated include autoimmune
CC diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g.
CC neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac
CC arrest, cerebrovascular disorders e.g. cerebral ischaemia, angiogenesis,
CC nervous system disorders e.g. Alzheimer's disease, infections caused by
CC bacteria, viruses and fungi and ocular disorders e.g. corneal infection,
CC and many other disorders listed in the specification. The polypeptides
CC can also be used to aid wound healing and epithelial cell proliferation,
CC to prevent skin aging due to sunburn, to maintain organs before
CC transplantation, for supporting cell culture of primary tissues, to
CC regenerate tissues and in chemotaxis. The polypeptides can also be used
CC as a food additive or preservative to increase or decrease storage
CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,
CC minerals, cofactors and other nutritional components. The present
CC sequence represents a novel secreted protein of the invention. Note: The
CC sequence data for this patent did not form part of the printed

Query Match 5.7%; Score 582; DB 4; Length 134;
Best Local Similarity 91.9%; Pred. No. 2.8e-30;
Matches 114; Conservative 2; Mismatches 8; Indels 0; Gaps 0;
QY 1831 KRGNDLCKRKMEENTSNLSKQESFTSVKPKRDSKDLALCSMLTETMETHEDAWPFL 1890
Db 1 KRGNDLCKRKMEENTSNLSKQESFTSVKPKRDSKDLALCSMLTETMETHEDAWPFL 60
QY 1891 LPVNLKLVPGYKKVKKPKMDFSTIREKLSGSGYPNLETALDVLRLVFNCFETFNEDDSDI 1950
Db 61 LPVNLKLVPGYKKVKKPKMDFSTIREKLSXGQYPNLETALXVRLVFDTXETFMCKXXDI 120
QY 1951 GRAG 1954
Db 121 GRAG 124

RESULT 35
ABUS5694

ID ABUS5694 standard; protein; 134 AA.
XX AC ABUS5694;
XX DT 18-MAR-2003 (first entry)
XX DE Human novel polypeptide #781.
XX KW Human; neural disorder; immune system disorder; renal disorder;
KW muscular disorder; respiratory disease; reproductive disorder;
KW gastrointestinal disorder; pulmonary disorder; cardiovascular disorder;
KW hyperproliferative disorder; inflammatory disease; allergic reaction;
KW blood related disorder; cancer; immunosuppressive; antiinflammatory;
KW cardiovascular; nephrotropic; cytostatic; antiallergic; thrombolytic;
KW haemostatic; antiarteriosclerotic.
XX OS Homo sapiens.
XX PN US2002132753-A1.
XX PD 19-SEP-2002.
XX PF 17-JAN-2001; 2001US-00764864.
XX PR 31-JAN-2000; 2000US-0179063P.
PR 04-FEB-2000; 2000US-0180628P.
PR 28-JUN-2000; 2000US-0214886P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216800P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225475P.
PR 14-AUG-2000; 2000US-0225758P.
PR 22-AUG-2000; 2000US-0226868P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 08-SEP-2000; 2000US-0231413P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 27-SEP-2000; 2000US-0235834P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241809P.
PR 01-NOV-2000; 2000US-0244617P.
PR 17-NOV-2000; 2000US-0249299P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.

```

PR 08-DEC-2000; 2000US-0251869P.
XX
PA (ROSE/) ROSEN C A.
PA (RUBE/) RUBEN S M.
PA (BARA/) BARASH S C.
XX
PI Rosen CA, Ruben SM, Barash SC;
XX
XX WPI; 2003-147444/14.
DR N-PSDB; ABX73953.
DR
XX
XX New polypeptides and nucleic acids, useful in gene therapy for treating,
PT inhibiting or preventing e.g. neural, immune system, muscular,
PT respiratory, reproductive, gastrointestinal, pulmonary, cardiovascular or
PT renal disorders.
XX
XX Claim 11; SEQ ID NO 1578; 402pp; English.
XX
XX The invention relates to human novel polypeptides and their associated
CC polynucleotides. The polypeptides and polynucleotides are useful in gene
CC therapy for treating, inhibiting or preventing neural disorders, immune
CC system disorders (e.g. systemic lupus erythematosus, rheumatoid arthritis
CC and multiple sclerosis), muscular disorders, respiratory diseases (e.g.
CC nasal vestibulitis, nasal polyps and sinusitis), reproductive disorders,
CC gastrointestinal disorders, pulmonary disorders, cardiovascular disorders
CC (e.g. congenital heart defects, Ebstein's anomaly and hypoplastic left
CC heart syndrome), renal disorders (e.g. acute kidney failure and end-stage
CC renal disease), hyperproliferative disorders (e.g. Hodgkin's disease and
CC leukaemia), inflammatory diseases (e.g. septic shock, burstsitis and
CC appendicitis), allergic reactions and conditions (e.g. asthma), blood
CC related disorders (e.g. thrombosis, atherosclerosis and myocardial
CC infarction) and cancerous diseases. Sequences ABUS4914-ABUS5699 and
CC ABUS5748 represent human novel polypeptides of the invention
XX
XX Sequence 134 AA;
XX
XX Query Match 5.7%; Score 582; DB 6; Length 134;
XX Best Local Similarity 91.9%; Pred. No. 2.8e-30;
XX Matches 114; Conservative 2; Mismatches 8; Indels 0; Gaps 0;
XX
QY 1831 KRGKNDLQKRMEENTSINLSKQESFTSVKPKRDSKDLCMSILTEMETHEDAWPFL 1890
DB 1 KRGKNDLQKRMEENTSINLSKQESFTSVKPKRDSKDLCMSILTEMETHEDAWPFL 60
XX
QY 1891 LPVNLKLVPGYKKVJIKKPMDFSTIREKLSXGQYPNLETFALDVRVFNCFTHEDDDSDI 1950
DB 61 LPVNLKLVPGYKKVJIKKPMDFSTIREKLSXGQYPNLETFALXKRVLFVDTXETFMKXXDI 120
XX
QY 1951 GRAG 1954
DB 121 GRAG 124
XX
RESULT 36
AAB95562
ID AAB95562 standard; protein; 779 AA.
XX
AC AAB95562;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human protein sequence SEQ ID NO:18199.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
XX 07-FEB-2001.
PD
PF 28-JUL-2000; 2000EP-00116126.
XX

```

```

PR 29-JUL-1999; 99JP-00248036.
PR 27-AUG-1999; 99JP-00300253.
PR 11-JAN-2000; 2000JP-00118776.
PR 02-MAY-2000; 2000JP-00183767.
PR 09-JUN-2000; 2000JP-00241899.
XX
XX (HELI-) HELIX RES INST.
XX
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
XX WPI; 2001-318749/34.
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602 full-
PT length cDNAs defined in the specification, and for the detection and/or
PT diagnosis of the abnormality of the proteins encoded by the full-length
PT cDNAs.
XX
XX Claim 8; SEQ ID NO 18199; 2537pp + Sequence Listing; English.
XX
XX The present invention describes primer sets for synthesising 5602 full-
CC length cDNAs defined in the specification. Where a primer set comprises:
CC (a) an oligo-dT primer and an oligonucleotide complementary to the
CC complementary strand of a polynucleotide which comprises one of the 5602
CC nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in the
CC specification. The primer sets can be used in antisense therapy and in
CC gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH1633 to AAH18742 represent human cDNA sequences; AAB2446 to AAB95893
CC represent human amino acid sequences; and AAH13629 to AAH1632 represent
CC polynucleotides, all of which are used in the exemplification of the
CC present invention
XX
XX Sequence 779 AA;
XX
XX Query Match 5.5%; Score 558; DB 4; Length 779;
XX Best Local Similarity 25.3%; Pred. No. 1.6e-27;
XX Matches 215; Conservative 130; Mismatches 292; Indels 212; Gaps 29;
XX
QY 46 SSNSD---SDSGTSSDTSSEGI-----SSSDSDLEDEDEEDQSIERESDDSDSEEA 97
DB 46 SLNGDVNVNGLSTVSHTTTSGILNSAPHSSSTLHHPSVAYDCLWNYSQPSANPGSNL 105
XX
QY 98 QHK-----SNNQVLLHGI-----SDPKADGQKATEKAEKRI-----HQPLPLAFESQT- 141
DB 106 KDPPLLQFSGQYPLNGILGSGRQSPSPSHNTNLRAGSQEFWANGTQSPNGLNFDSDQL 165
XX
QY 142 -HSFQSQKQPVLSQQLPFIFQSSQAKESVNNKHTSVIQTGLVSNVKPISLVNQAKKE 200
DB 166 YDSPFDQNF--EVMFNGPPSPFTSPQSPMLGSSITQTFAPSQEVGSGIHP---DEAAEKE 220
XX
QY 201 TYMKLIYPSDVLKAGNKNTSESSLLTSELRSKREQ-----YKQAFPQLKKQESSK 253
DB 221 --MTSVV-----AENGTGLVGSLELEEEQELKCMGVNGSVP-----SVE 258
XX
QY 254 SLKKVIAALSNPKATSSPAHPKQTLNHNHPNFLTALLGNHQPNGVIOQVIOEAPLAL 313
DB 259 SLHGEVSVLV-PDPTVSCLDLP-----SHLPD-----QLEDTPI-- 291
XX
QY 314 TTKTKMQSKINENIAAASSTPFSPVNLSTSGRTPGNTQTPVMPSPASPIHLSQGKEKAVS 373
DB 292 -----LSEDSLEPFNSLAPEVPVSGGLYGIDDTTELM-----GAEDKLP 328
XX

```


Db 615 EFLSCYGLLLPDAQYPIITAVSLMEALSADKGGFL-YLNRVLVILLQTLQDE-IAEDY- 671

Qy 962 AKTALGEHLNNGVNRDNNVSEILQFM-----BAHCOQTELT 998

Db 672 --GELGMLKSEIPLTLHVSSELVRLUCRRSDVQEESEGSDTDNDKSDAAFDNEVDDEFL 729

Qy 999 ESLTKAFQAHTPAQKASVLAFLINELACSKSVVSEIDKNIDYMSNLRDRKVV-----EG 1054

Db 730 EKLETSFELTSEKQLQILATLCHRILMTYSVDHMETRQMSAELWKERLAVLKEEND 789

Qy 1055 KLRKLRIIHAKTGTRDGGIDGEOHPGLGTPPGRRRRKRGSGSDYDDDDDDSD- 1113

Db 790 KRAEKQRKXEMAKNGKVGNG-----LG-----KTRKKEIVKFPQVDTEAEDM 838

Qy 1114 -----QGEDDEDEEDKQKKTDCEDDEGDDQASVEELEKQIEKLSQ 1161

Db 839 ISAVKSRLLAIQAKEREIOE--REMVKVLERQAEERIRKHAAEK----- 885

Qy 1162 OSQYRKLFDASHLSRSMFGGDRVRRRYWIL-PRCGGIFVEGMSGEGLEIAKEREKL 1220

Db 886 --AFQEGIAKAKLVNRTPIGTDRHNRYWLFSDVPLGFI----- 925

Qy 1221 KKAESVQIKEMFTSG---DSLNCSTHCEQKEDLKEKONTNLFLOKPGSFSLKSL 1277

Db 926 -----KGWVHDSIDYRFNHHC-----KQHT---VSGDEDYCPRSK-- 957

Qy 1278 EVAKMPPSEVMTPKPNAGANGCTLSYQNSGKHSLSGVSTATQSNVEKADSNLFTG 1337

Db 958 -----KANIGKNA-----SMNTOHGTATEVAVE----- 980

Qy 1338 SGPGRFYSPNPLQLLKTLTEKQ--WFSLLPRTPCDDTSLTHADMTASLVTQSQPP 1395

Db 981 -----TTTPKQGNLNL-----CD----- 995

Qy 1396 SKSPSPTPAPLGSSAQNPVGLNPFALSPLOKGVMMGLQFCGWPTGVWTSNIPFTLSV 1455

Db 996 ----- 995

Qy 1456 PSLGSLGLSEGNNGSFLTSNVASSKSPVPONEKATSAQPAVEVAKPVDPPSPKIP 1515

Db 996 ----- 995

Qy 1516 BEMQFQWRIIDPEDKALLKVLHURGIREKALOKIOK-HLDYITOACLKNN-DVAIE 1573

Db 996 -----SQKDELNCLHPOGIRESQLERLEKRYQDIHSLHARKNPLGLKS 1044

Qy 1574 LNNEE--NOVTRDIVNWSVEEQAMDLVLOQVEDLRRVAS-----ASLQ 1620

Db 1045 CDGNQELNLFRLSLIE---VATRLQKGLGTYEETSFEARVISLEKLDKDFGECVIALQ 1101

Qy 1621 -----VKGWMCPPASEREDLVYFEHKSFTKLCKEHDGFTGDESSAHALERKSDNP 1673

Db 1102 ASVIKKFLQGFMAPQO-----KRRKLQS 1124

Qy 1674 LDIAVTRLADLERNTERRIEEDIAPLVRWRALSEARSAAQVLCIOLOKSIAXEKSI 1733

Db 1125 EDSATEEVEDEKQVE--EAKVAGALEKWKTAIREAOTFSRHVLLGLMDACIKWDNSA 1182

Qy 1734 MKVYQICIRKGNBELLILCDGCDKCHTYCHRPKITTIPODGNFCPACIAKASQTLKI 1793

Db 1183 ENARCKVCKKGEDDKILDCBNKAFHLFCLRLPALYEVDPGEWQCPACQAPATA----- 1236

Qy 1794 KKLHVKKTNESKGGKVTLTGDTEDDS----- 1823

Db 1237 --RRNSRGNYTBESASE-----DSEDDSESEEEEEEEEDYEVAGLRPRKTIR 1289

Qy 1824 -----ASTSSLRKGNKDLQKRWEENTS--INLSQESFTSVKPKRDDSDDLALCS 1874

Db 1290 GKHSVIPPAARSGRPGKPHSTRSQKAPPVDDAEVDELVLQTRSRRSQSLQKCE 1349

Qy 1875 MILTEMETHDAWPELLPNVLKLVPGYKVKIKKPMDFSTIREKLSGGQYNLETALDVR 1934

Db 1350 EILHKIVKYSWPPREPVRDEADYDVI THPMDFTQVQNKCSGYSRVSQEFPLDMK 1409

Qy 1935 LVFDCNCTFN 1944

Db 1410 QVFTNAEVYN 1419

RESULT 38

ADF69140

ID ADF69140 standard; protein; 1483 AA.

AC ADF69140;

XX 12-FEB-2004 (first entry)

XX Human MP53 protein sequence SEQ ID NO:110.

XX p53 pathway modulating agent; MP53; p53 modulator; cytostatic;

KW gene therapy; cancer; human.

XX Homo sapiens.

XX WO2003083047-A2.

XX 09-OCT-2003.

XX 28-FEB-2003; 2003WO-US006025.

XX 01-MAR-2002; 2002US-0361196P.

XX (EXEL-) EXELIXIS INC.

PI Belvin M, Francis-Lang H, Friedman L, Plowman GD, Heuer TS, Li D;

PI Funke RP;

XX WPI; 2003-812540/76.

DR N-PSDB; ADF69196.

XX Identifying a candidate p53 pathway modulating agent for treating e.g.,

PT cancer by contacting an assay system comprising a MP53 polypeptide or

PT nucleic acid with a test agent and detecting a test agent-biased

PT activity.

XX Example; SEQ ID NO 110; 406pp; English.

PS The present invention describes a method for identifying a candidate p53

CC pathway modulating agent, which comprises: (a) providing an assay system

CC comprising a MP53 (modulator of p53) polypeptide or nucleic acid or its

CC fragment or derivative; (b) contacting the assay system with a test agent

CC under conditions where the system provides a reference activity except in

CC the presence of the test agent; and (c) detecting a test agent-biased

CC activity, where a difference between the test agent-biased activity and

CC the reference activity identifies the test agent as a candidate p53

CC pathway modulating agent. Also described: (1) modulating the p53 pathway

CC of a cell; (2) modulating the p53 pathway in a mammalian cell; and (3)

CC diagnosing a disease in a patient. MP53 has cytostatic activity, and can

CC be used in gene therapy. The method is useful for identifying a candidate

CC p53 pathway modulating agent for preparing a composition for diagnosing

CC or treating e.g., cancer. The present sequence represents a human MP53

CC protein, which is used in the exemplification of the present invention.

XX SQ Sequence 1483 AA;

Query Match 5.3%; Score 533; DB 7; Length 1483;

Best Local Similarity 18.0%; Pred. No. 1.9e-25;

Matches 348; Conservative 253; Mismatches 585; Indels 744; Gaps 66;

Qy 221 SESSSLTSETSRKREQY-----KQAFPSQLKKQESSKSLKKVIAALSNNPKATSSSPA 273

Db 28 TQEAFTREYEARELRSERIWTCKSTGSQLTHKEAWEEBEEQVAVELK-----EETPA 82

Qy 274 -HPKQTLNHNPNFLTNALLGNHQPNGVIOEQAPLALTTKTKMOSKINENIAAASS 332

Db 83 WYKLVLEMHVN-----TASLEKLVDTAWLEIMTKYAVGECDEFEV----- 124

PF 17-APR-1998; 98WO-JP001783.
XX
PR 18-APR-1997; 97JP-00116570.
PR 24-OCT-1997; 97JP-00310027.
XX
PA (CHUS) CHUGAI RES INST MOLECULAR MEDICINE INC.
XX
PI Jones MH;
XX
XX WPI; 1998-583603/49.
DR N-PSDB; AAV68409.
XX
PT Transcriptional regulator gene family containing bromodomain - may be
PT expressed in testis tissue and is useful for treatment of cancer and
PT other proliferative disorders.
XX
XX Claim 1; Page 145-157; 187pp; Japanese.
PS
XX This sequence represents the human BAZ1-beta protein, a member of a
CC family of transcriptional regulator genes containing a bromodomain (BAZ,
CC Bromodomain with Atypical Zinc Finger) which are expressed specifically
CC in testis tissue and also in certain tumour lines. Transgenic cells may
CC be used for the preparation of the BAZ1-alpha, BAZ1-beta, BAZ2-alpha and
CC BAZ2-beta proteins. These proteins can be used in the treatment of cancer
CC and other proliferative disorders, and in screening of compounds for
CC their binding ability to the expression products (e.g. for use as drugs
CC by modulation of transcriptional regulation)
XX
XX Sequence 1531 AA;
Query Match 5.3%; Score 533; DB 2; Length 1531;
Best Local Similarity 17.9%; Pred No. 2e-25;
Matches 352; Conservative 256; Mismatches 600; Indels 756; Gaps 67;
QY 189 KPLSLVNOAKETVMKLI VSPDVLKAGNKNTSBESSLLTSELRSKREQY-----KQA 241
Db 8 KPFPVLNLPGEPP-FTIPH-----TQAPRTREYEARELERYSERIWTCKST 55
QY 242 FPSQLKKQESKSLKVVIAALSNPKATSSSPA-HPKQTLNHNHPFLTNALLGNHQPNG 300
Db 56 GSSQLTHKEAWEEQEVAAELK-----EFPAYVEKLVLEMVHN-----TA 97
QY 301 VIQSVIOEAPLALTTKTQWQKINENIAAASSTPSPVNLSTSGRRTPGNGTPWMPAS 360
Db 98 SLEKLVDTAWLEIMTKYAVGECDEV----- 124
QY 361 PILHSQGEKAVSNVNPVKQHHSHPAKSLVEQFRGTDSDIPSSKSDSENEDEEDE 420
Db 125 -----GKEKML--KVKIVKI-----HPL-----EKVDEE 146
QY 421 EDEDEDDDDSDQSSESDNSDTEGSEEDDDDDXQDESDD-----T 467
Db 147 ATEKSDGACDSPSSDKSNSQIAQDHQKXETTVKVEDGRRESINDRARRSPKLPSTSLK 206
QY 468 EGE-----KTSMKL--NKTSSVKSPSLGTHSTPRLNLIHAK----- 503
Db 207 KGERKWPAPKPLPHKYDVLQWEDKLIISNV--PADSLIRTPPPNKEIVYFIRHNALRA 264
QY 504 -----APGSAAPALCSESQSPA-----FLGTSSSTL-----TSSPHSGTSKRRRTV 544
Db 265 GTGENAPVWVEDELVKKSLSPSKFSDFLLDPYKYMTLNPSTKRNKVTGSPDRKPSKSK-T 323
QY 545 DERELRIPLFVG-WORETRIRNFGG---RLQGEVAYAPCGKLLROYPEVIKYLNRGIM 600
Db 324 DNSLSLSPNPKLWCHVHLKSLSGSLKSPKVNKSNKSP-----BEHLEEMKMMSPN--- 376
QY 601 DISRDNFSFSAKIRVGDVEARDGQPOEMQWCLLKEDVIPRIAMEGRGRPPNPDRORA 660
Db 377 -----KLHT--NFHIPKKG----- 395
QY 661 REESMRRRKGRPPNVGNAEFLDNADAKLLRKLQAEIARQAAQIKLRLKLOQOARVA 720
Db 396 KHSKPLKAGRSKGI-----LNGQKSTGNS 421

QY 721 KEAKKQQAIAAAEERKQKEQIKIMQOEKIKRIQQ-----IRMEKEL--RAQOIL 769
Db 422 KSPKK--GLTKPKTKMKQMTLLDMAKGTQKTRAPRNSGGTPRTSSSKPHKILPPAALHLI 479
QY 770 EAKKKKEE-----AANAKLLEAEKRIK-EKEMR---RQAAVLLKHQER----- 809
Db 480 AYYKENKDRKRSALSCVISKTARLLSSSEDRARLPEELRSLVQKRYELLEHKRWASMS 539
QY 810 ERRQHMLMKAMEARKAEBKERLKOBRDEKLNKRLKLEQRLELEMAKELKPNED 869
Db 540 EQRKEYLKKREBELKKLKEKAKERKEKLEKQRYEDQEL---TGKNL----- 590
QY 870 MCLADQQLPLPRLPRIGLVLSGSPFSCLMVVOQLRNFQKVLGVDVNI DVENLSVLOEGL 929
Db 591 -----PAFLVDTPB-GLPNTLFGDVAMVVEFLSCYSGLLLPDAQYPIITAVSLEALS 642
QY 930 LNI GD--SMGEVQDLLVRLLSAAVCDPLITGYKAKTALGEHLNVGNRDNVSEILOIF 987
Db 643 ADKGFLYLRNVILLQTLQLLQDEIAEDY---GELGMKLSIPLTLHSVSELVRLC 699
QY 988 M-----EAHCGQTELTESLTKAFOAHTPAQKASVLAFLINE 1024
Db 700 LRRSDVOEESGSDTDNKKDSAAFEDNEVQDEFLEKLETSFEFFELTSEKLIQTALCHR 759
1025 LACSKSVVSEIDKNIDYMSNLRDRKWV---EGKLRLRIIHAHKTGKRTDTSGGIDIGE 1080
Db 760 ILMTYSVDHMETQOMSALWKEKRLAVLKEENDKRAEKQKREMEAKNKGKVENG- 818
QY 1081 EQHPLGTPTPGKRKRKRGSDSDYDDDDDDDDDD-----QDEDEDEDEEDKED 1127
Db 819 ---LG-----KTDRKKRIVKFPQVQVDEAEDMISAVKSRRLALAIQAKKEREIQE--RE 866
QY 1128 QKGKTDICEDDEGDGQAASVEELEKQIEKLSQOQSVRRKLPDASHLSRSMFGPDYR 1187
Db 867 MKVLERQAEERIRKHAAAEK-----AFQEGIAKAKLVMRRTPIGTDRNH 913
QY 1188 RRYWIL-PRCGGIVFEGMESGELEETAKERELKKAESVQIKEEMFTSG---DSLACS 1243
Db 914 NRYWLFSDVPLGLFIE-----KGVHDSIDYR 940
1244 NTDHCEQKEDLKEKDNLTNLFQKPGSPSKLSKLEAVAMPPESEVMTPKPNAGANGCTLS 1303
Db 941 FNHHC-----KDHT---VSGDEDYCPRSK-----KANLGKNA----- 969
QY 1304 YQNSGKSHSLGVSQSTATQSNVEKADSNLFWTSGSGPGKFYSPLPNDQLLKTLEKTRQ- 1362
Db 970 -----SMNTQHGTA TEVAVE-----TTTPKQGN 993
QY 1363 -WFSLLPRTPCDDTSLTHADMSLASLVTPOQSPSPSPPTPAPLGSSAQNVPGLNPPAL 1421
Db 994 LWFL-----CD----- 999
QY 1422 SPLQVKGVSMMQLQFCGWTGVVTSNIPFSLVPSLGSGLGSEGNGNSFLTNSVASSK 1481
Db 1000 ----- 999
QY 1482 SESVPQNEKATSAQAAVEVAKVPDFSPKPIPEEMQFGWRIIDPDLKALLKVLHLR 1541
Db 1000 -----SKELDELLNCLHPQ 1014
QY 1542 GIREKALQKOIQ-HLDYITOACLKNK-DVAIIENENEE--NQVTRDIVENWSVEEQAM 1597
Db 1015 GIRESQLERLEKRYQDIHSHLARKPNGLKSCDGNQELLNLRSLIE---VATRLQ 1071
QY 1598 EMDLSVLQVQEDLERRVAS-----ASLQ-----VKGMWCPPEASEREDLVYF 1639
Db 1072 KGGLYVEETSEFARVVISLEKLDGFCVETALQASVIKKFLQGFMAKPKQ----- 1121
QY 1640 EHSFTKLCHEHDGEFTGEDESSAHLERKSDNPLDIAVTRLADLERNIERRIEEDIA 1699
Db 1122 -----KRRKLOSEDSAKTEEVDEEKKWVE--EAKVASA 1152

QY 1700 LRVWRALSEARSAQAVALCQLOKSIWAKSIAWYQICRKGDNBEILLILCDGCKG 1759
DB 1153 LEKWTAREATFRSMHYLLGLMDACIKWMSAENARCKVCPKRGEDDKLILCDECNKA 1212
QY 1760 CHTYCHRPKITIPGDWFCPACIAKASGOTLKIKLHVKGKKTNESKKGKVTLTGDT 1819
DB 1213 FHLFCLRPALYVPGWQCPACQAPATA-----RNSRGRNYTEBSASE-----JSE 1259
QY 1820 DEDS-----ASTSSILKRGNKDQLQKRM 1842
DB 1260 DDESDEEBEEBEEDYEAGLRRLPRKTIIRGHKSVIPPAARSGRRPGKPHSTRRS 1319
QY 1843 BENTS--INLSQESFTSVKPKRDSKDLALCSMLTETHEADWPFLPVNKLVP 1900
DB 1320 QKAPPPVDAEVLQTKRSSRSQSLQKCEILHKIVKIRFSWPFREPVTREABD 1379
QY 1901 YKVIKPKMDFITREKLSGGYPNLETFALDVRVFDNCETFN 1944
DB 1380 YDVIHPMDFQTVQNKSCGYSRVOEFLTMKQVFTNAEYVN 1423

RESULT 40

AAW81172
ID AAW81172 standard; protein; 1527 AA.

AC AAW81172;

DT 05-MAY-1999 (first entry)

DE Human BAZ1-beta protein #1.

KW Transcriptional regulator; BAZ1-alpha; bromodomain; BAZ;
KW atypical zinc finger; testis; human; tumour; BAZ1-beta; BAZ2-alpha; drug;
KW BAZ2-beta; treatment; cancer; proliferative disorder; screening.

OS Homo sapiens.

PN WO9847920-A1.

PD 29-OCT-1998.

PF 17-APR-1998; 98WO-JP001783.

PR 18-APR-1997; 97JP-00116570.

PR 24-OCT-1997; 97JP-00310027.

PA (CHUS) CHUGAI RES INST MOLECULAR MEDICINE INC.

PI Jones MH;

DR WPI; 1998-583603/49.

DR N-PSDB; AAV68408.

PT Transcriptional regulator gene family containing bromodomain - may be
PT expressed in testis tissue and is useful for treatment of cancer and
PT other proliferative disorders.

PS Claim 1; Page 125-137; 187pp; Japanese.

XX This sequence represents the human BAZ1-beta protein, a member of a
CC family of transcriptional regulator genes containing a bromodomain (BAZ,
CC Bromodomain with Atypical Zinc finger) which are expressed specifically
CC in testis tissue and also in certain tumour lines. Transgenic cells may
CC be used for the preparation of the BAZ1-alpha, BAZ1-beta, BAZ2-alpha and
CC BAZ2-beta proteins. These proteins can be used in the treatment of cancer
CC and other proliferative disorders, and in screening of compounds for
CC their binding ability to the expression products (e.g. for use as drugs
CC by modulation of transcriptional regulation)

XX Sequence 1527 AA;

Query Match 5.2%; Score 532; DB 2; Length 1527;
Best Local Similarity 18.0%; Pred. No. 2.3e-25;

Matches 354; Conservative 255; Mismatches 597; Indels 756; Gaps 68;
QY 189 KPLSLVNOAKETYMKLVPSDVLKAGNKNTSESSLTSLRSKREOY-----KOA 241
DB 8 KPFLVNPPLPGEPEF-FIIPH-----TQEAFTREEYEARLERYSERIMWCKST 55
QY 242 FPSQVLKQESSKSLKKVYAAALSNPKATSSSPA-HPKQTLNHHNPFLTNALLGNHQPNG 300
DB 56 GSSQLTHKEAWEEEOEVAELLK-----EFPAPWYEKLVEVMVHN-----TA 97
QY 301 VTQSVIQAPIALTTKTKMQSKINENIAAASSTPFSSPVNLTSGRRTPGNQTPVMPSPAS 360
DB 98 SLEKLVDFANLEIMTYAVGEECDFEV-----124
QY 361 FILHSQSGKEKAVNNVNPVKTOHHGHPAKSLVQEFRGTDSDIPSSKSDSEDEEEDDE 420
DB 125 -----GKEKML--KVKIVKI--HPL-----EKVDEE 146
QY 421 BEDEEDDDDDSDSQSSDSNSSESDTSGSEEDDDDDKQDSDSD-----T 467
DB 147 ATEKSGDAGCDSPSSDKENSSQIADHQKKTETVVKEDGGRRESINDRARRSPKLPSTSLK 206
QY 468 EGE-----KTSMKL---NKTTSVKSPSLTGHSTPRNLHIK-----503
DB 207 KGERKWAPPKFLPHKYDVKLQNEDKIISNV--PADSLINTERPPNKEIVRYFIRHNALRA 264
QY 504 -----APGSAPAAALCSESQSPA-----FLGTSSSTL---TSSPHSGTSKRRRV 544
DB 265 GTGENAPWVVEDELVKYSLPSKFSDFLLDPKYMTLNPSTVKRYGTSDDPKPSKSK-T 323
QY 545 DERELRIPIEYV-GWORETRIIRNFGG---RLQGEVAYAPCGKKLQYPPVIKYSNNGIM 600
DB 324 DNSLSLSPNPKLMCHVHLKKSLSGSLPKVKNKSKSP-----EEHLEEMKMMSPN--- 376
QY 601 DISRDNFSFAKIRVDFEYARDGQEMQWCLLKEDVPIPRAMEGRGRPNPDQRA 660
DB 377 -----KLHT-NFHI PKKGP-----PAKKGP 395
QY 661 REESRRRRKGRPNVGNNAEFLDNADAKLLRKLQAEIARQAQAIKLLRKLQKQEARVA 720
DB 396 KHSDFPLAKGSKGI-----LNGQSKTGN 421
QY 721 KEAKQQAQIAAEKREKQEKIKMKQOEKIKRIOO-----IRMEKEL--RAQQL 769
DB 422 KSPKK--GLKTPKTKQMTLLDMAKGTQKWTAPRNSGGTPTRTSSKPHKHLPPAALHLI 479
QY 770 EAKKKKEE-----AANAKLLEAEKRIK-EKEMR---RQAVLLKHQER-----809
DB 480 AYYKENKDRBKRSALSCVISKARLLSSEDRARLPSELRLSVQRYELLEHKRWASMS 539
QY 810 ERRROHMLMKAMEARKKAEKERLQEKREKRLNKKERKLBORLELEMAKELKKPNED 869
DB 540 EQRKEYLKKREELKKLKEKAKERREKEMLERLEKQKRYEDEL---TGKVL-----590
QY 870 MCLADQKPELPRIPLGLVLSGSTFSDCLMVVQFLRNFGKVLGPDVNDVPSLVSLEQGL 929
DB 591 -----PAFRLVDTFE-GLPNTLPGDVAMVVFELSCYSGLLPDAQYPIITAVSLMEALS 642
QY 930 LNIQDSMEVQDILLVRLLSAAVCDPLITGYKAKTALGEHLLNVGNVRNDSVSEIIQIFM- 988
DB 643 ADKGGFL-YLNRVLVILLQTLQDE-IAEDY---GELGMKLESEIPLTLHSVSELVRLCIUR 697
QY 989 -----EAHCGOTELTESLTKTAFOAHTPAQKASVLAFLINELA 1026
DB 698 RSDVQEESEGGSDTDNKDSAAFEDNEVQDEFLEKLETSEFFELTSEKLIQIITALCHRL 757
QY 1027 CSKSVYSIDKNIDYMSNLRDKWVY---EGLRKLRIIHAKTGKRDTSGGIDLGEBO 1082
DB 758 MTYSVQDHMETRQMSAELMWERLAVLKEENDKKAQKQKMEAKNKENGKVENG---814
QY 1083 HPLGTPTPGRKRRKGGSDYDDDDDDSD-----QGEDDEDEEEDKEDOK 1129
DB 815 --LG-----KTRDKRIVKFPQVDTAEADMISAVKSRRLAIQAKKEREQE--REMK 864

```
Qy 1130 GKTTDICEDEGQAAVEELEKQIEKLSKQOQOYRRLKFDASHSLRSVMFGPDYRERR 1189
Db 865 VLEROAEERIRKHAAEK-----AFQEGIAKAKLVMRRTPIGTDRNHR 911
Qy 1190 YWIL-PRCGGIFVEGMESGEGLEETAKEREKLLKKAESVQIKEEMPETSG----DSLNCST 1245
Db 912 YWLFSDVPGLFIE-----KGWVHDSIDYRFN 938
Qy 1246 DHCEQEDLKEDNDTNLFLQKPGSFKLSKLELVAKMPPSEVMTPKPNAGANGCTLSYQ 1305
Db 939 HHC-----KDHT--VSGDEDYCPRSK-----KANLGKNA----- 965
Qy 1306 NSGRHSLGVSOSTATOSNVEKADSNLNTGSSGPGKPYSPLPNDQLLKTITKRNQ--W 1363
Db 966 -----SMTQGTATEVAE-----TTTTQGGQNLW 991
Qy 1364 FSLLPRTCPDTSLTADMTASLVTQSPQSPKSPPTAPLGSAAQNPVGLNPFALSP 1423
Db 992 FL-----CD----- 995
Qy 1424 LQVKGVSMMGLQFCGWTGVTWTSNIPFTLSVPSLGSGLGSLSENGNSFLTNSVASSKSE 1483
Db 996 ----- 995
Qy 1484 SPVPQNEKATSAQPAAEVAKVPDPFPPKPIPEEMQFCGWRRIIDPEDIKALLKVLHLRGI 1543
Db 996 -----SQELDELLNCLHPOGI 1012
Qy 1544 REKALQKQIQK-HLDYITQACIKNK-DVAIIELNENE--NQVTRDIVENMSVBEQAMEM 1599
Db 1013 RESQLKERKRYQDIHSHILAKPNLGLKSCDGNQELLNFRSLDIE---VATRLQKG 1069
Qy 1600 DLSVLQQVEDLERRVAS-----ASLQ-----VKGWMCPEPASREDLVYFEH 1641
Db 1070 GLGYVEETSEFEARVISLEKLKDFGECVIALQASVIKKFLQGFMAKQ----- 1117
Qy 1642 KSFTKLCHEGDEFTGEDESSNAHALERKSDNPLDIATVTRLADLERNIERRIEEDIAPCLR 1701
Db 1118 -----KRRLOSEDSAKTEEVDEEKKWVE--EAKVASALE 1150
Qy 1702 VMRRALSEARAAQVALCIIQIQKSIKQIAWEKSIKMYVCOICRKGDNNEELLILLCDGCKGCH 1761
Db 1151 KWKTAIRAQAQFISRMHVLGLMLDACIKWMSAENARCKVCPKCGEDDKLILCDECNKAFH 1210
Qy 1762 TYCHRPKTTIPDGNWFCPACIAKASGOTLKIKLHVKGKTNESKGGKVTTLFGDTDEDE 1821
Db 1211 LFLCLRALYEVDPGEWQCPACQAPATA-----RNSRGRNYTEESASE-----DSEDD 1257
Qy 1822 DS-----ASTSSSLKGNKDLQKRKWE 1844
Db 1258 ESEEEEEEEEEEEEDVEVAGLRPRKRTIRGKHSVTPPAARSGRRFGKPKHSTRSQP 1317
Qy 1845 NTS--INLSKQESFTSVKPKRDDSOLALCSMILTEMETHEDAWPFLLPVNLKLVPGYK 1902
Db 1318 KAPPVDDAEVDELVLQTKRSRRQSLQKCEEILHKIVKYPFSWPFREPVRDEADYY 1377
Qy 1903 KVIKKPMDFTIREKSLSGQVNPNETFALDVLRLVFNDCETFN 1944
Db 1378 DVITHPMDFTQVQKSCGSYSRVQEFLLTMKQVFTNAEVYN 1419

RESULT 41
ABB37253
ID ABB37253 standard; peptide; 97 AA.
XX AC ABB37253;
XX AC ABB37253;
DT 04-FEB-2002 (first entry)
DE Peptide #4759 encoded by human foetal liver single exon probe.
XX Human; foetal liver; gene expression; single exon nucleic acid probe.
KW
```

```
XX Homo sapiens.
OS WO200157277-A2.
PN 09-AUG-2001.
PD 30-JAN-2001; 2001WO-US0006669.
PF 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
PA Penn SG, Hanzel DK, Chen W, Rank DR;
PI WPI; 2001-483447/52.
XX Human genome-derived single exon nucleic acid probes useful for analyzing gene expression in human fetal liver.
PT Claim 27; SEQ ID NO 29888; 639pp + Sequence Listing; English.
PS The invention relates to a single exon nucleic acid probe for measuring human gene expression in a sample derived from human foetal liver. The single exon nucleic acid probes may be used for predicting, measuring and displaying gene expression in samples derived from human fetal liver. The present sequence is a peptide encoded by a single exon nucleic acid probe of the invention. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
SQ Sequence 97 AA;
Query Match 4.8%; Score 491; DB 4; Length 97;
Best Local Similarity 100.0%; Pred. No. 1.8e-24;
Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 439 SDSNSEDTEGSEEDDDDDKQDESDDSDTEGKTSMLKNTTSSVKSPSMLTGHSTPRN 498
Db 1 SDSNSEDTEGSEEDDDDDKQDESDDSDTEGKTSMLKNTTSSVKSPSMLTGHSTPRN 60
Qy 499 LHIKAPGSAAPALCSESQSPAFLTGTSSTLTSSPHS 535
Db 61 LHIKAPGSAAPALCSESQSPAFLTGTSSTLTSSPHS 97

RESULT 42
AAM30718
ID AAM30718 standard; protein; 97 AA.
XX AC AAM30718;
XX 17-OCT-2001 (first entry)
DE Peptide #4755 encoded by probe for measuring placental gene expression.
XX Probe; microarray; human; placenta; antenatal diagnosis;
KW genetic disorder.
XX Homo sapiens.
XX WO200157272-A2.
XX 09-AUG-2001.
XX 30-JAN-2001; 2001WO-US0006663.
XX
```


PR 04-FEB-2000; 2000US-0180312P.
 PR 26-MAY-2000; 2000US-0207456P.
 PR 30-JUN-2000; 2000US-00608408.
 PR 03-AUG-2000; 2000US-00632366.
 PR 21-SEP-2000; 2000US-0234687P.
 PR 27-SEP-2000; 2000US-0236359P.
 PR 04-OCT-2000; 2000GB-00024263.
 XX (MOLE-) MOLECULAR DYNAMICS INC.
 PA Penn SG, Hanzel DK, Chen W, Rank DR;
 XX WPI; 2001-488897/53.
 XX Human genome-derived single exon nucleic acid probes useful for analyzing
 PT gene expression in human placenta.
 XX Claim 27; SEQ ID NO 30987; 654pp; English.
 PS The present invention relates to single exon nucleic acid probes (SENP:
 CC see AA1315-AA15746). The present sequence is a peptide encoded by one
 CC such probe. The probes are useful for producing a microarray for
 CC predicting, measuring and displaying gene expression in samples derived
 CC from human placenta. The probes are useful for antenatal diagnosis of
 CC human genetic disorders
 XX Sequence 97 AA;
 SQ Query Match 4.8%; Score 491; DB 4; Length 97;
 Best Local Similarity 100.0%; Pred. No. 1.8e-24; Mismatches 0; Indels 0; Gaps 0;
 Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 439 SDSNESDEGSEEDDDDDKQDESDSDTEGKTSMLKNTTSSVKSPMSLTGHSTPRN 498
 Db 1 SDSNESDEGSEEDDDDDKQDESDSDTEGKTSMLKNTTSSVKSPMSLTGHSTPRN 60
 QY 499 LHIAPGSAAPALCSQSQAFLGTSSLTSSPHS 535
 Db 61 LHIAPGSAAPALCSQSQAFLGTSSLTSSPHS 97
 RESULT 43
 ABB32005
 ID ABB32005 standard; peptide; 97 AA.
 AC ABB32005;
 XX 01-FEB-2002 (first entry)
 DT Peptide #4656 encoded by breast cell single exon nucleic acid probe.
 DE Human; microarray; single exon probe; gene expression; breast; disease;
 KW cancer.
 KW Homo sapiens.
 OS WO200157271-A2.
 XX WO200157271-A2.
 XX 09-AUG-2001.
 XX 30-JAN-2001; 2001WO-US0000662.
 XX 04-FEB-2000; 2000US-0180312P.
 PR 26-MAY-2000; 2000US-0207456P.
 PR 30-JUN-2000; 2000US-00608408.
 PR 03-AUG-2000; 2000US-00632366.
 PR 21-SEP-2000; 2000US-0234687P.
 PR 27-SEP-2000; 2000US-0236359P.
 PR 04-OCT-2000; 2000GB-00024263.
 XX (MOLE-) MOLECULAR DYNAMICS INC.
 PA Penn SG, Hanzel DK, Chen W, Rank DR;
 XX WPI; 2001-488897/53.

XX WPI; 2001-496933/54.
 DR New spatially-addressable set of single exon nucleic acid probes, useful
 XX for measuring gene expression in sample derived from human breast,
 PT comprises number of single exon nucleic acid probes.
 XX Claim 27; SEQ ID NO 14973; 327pp + Sequence Listing; English.
 PS The invention relates to a spatially-addressable set of single exon
 CC nucleic acid probes for measuring gene expression in a sample derived
 CC from human breast and BT 474 cells. The method involves contacting the
 CC probes with a collection of detectably labelled nucleic acids derived
 CC from mRNA of human breast, and then measuring the label bound to each
 CC probe of the microarray. The probes are useful for verifying the
 CC expression of regions of genomic DNA predicted to encode proteins. They
 CC are useful for gene discovery, and for determining predisposition and/or
 CC prognosing breast disease. Gene expression analysis is useful for
 CC assessing the toxicity of chemical agents on cells. The microarray of
 CC this invention presents a far greater diversity of probes for measuring
 CC gene expression, with far less bias than expressed sequence tag
 CC microarrays. The method is suitable for rapid production of functional
 CC information from genomic sequence. The present sequence is a peptide
 CC encoded by a single exon nucleic acid probe of the invention. Note: The
 CC sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences
 XX Sequence 97 AA;
 SQ Query Match 4.8%; Score 491; DB 4; Length 97;
 Best Local Similarity 100.0%; Pred. No. 1.8e-24;
 Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 439 SDSNESDEGSEEDDDDDKQDESDSDTEGKTSMLKNTTSSVKSPMSLTGHSTPRN 498
 Db 1 SDSNESDEGSEEDDDDDKQDESDSDTEGKTSMLKNTTSSVKSPMSLTGHSTPRN 60
 QY 499 LHIAPGSAAPALCSQSQAFLGTSSLTSSPHS 535
 Db 61 LHIAPGSAAPALCSQSQAFLGTSSLTSSPHS 97
 RESULT 44
 ABB22549
 ID ABB22549 standard; protein; 97 AA.
 AC ABB22549;
 XX 23-JAN-2002 (first entry)
 DT Protein #4548 encoded by probe for measuring heart cell gene expression.
 DE Human; gene expression; heart; microarray; vascular system;
 KW cardiovascular disease; hypertension; cardiac arrhythmia;
 KW congenital heart disease.
 OS Homo sapiens.
 XX WO200157274-A2.
 XX 09-AUG-2001.
 XX 30-JAN-2001; 2001WO-US0000666.
 XX 04-FEB-2000; 2000US-0180312P.
 PR 26-MAY-2000; 2000US-0207456P.
 PR 30-JUN-2000; 2000US-00608408.
 PR 03-AUG-2000; 2000US-00632366.
 PR 21-SEP-2000; 2000US-0234687P.
 PR 27-SEP-2000; 2000US-0236359P.
 PR 04-OCT-2000; 2000GB-00024263.
 XX (MOLE-) MOLECULAR DYNAMICS INC.
 PA Penn SG, Hanzel DK, Chen W, Rank DR;
 XX WPI; 2001-488897/53.

PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX Penn SG, Hanzel DK, Chen W, Rank DR;
 PI WPI; 2001-48899/53.
 XX
 XX Single exon nucleic acid probes for analyzing gene expression in human
 PT hearts.
 PT
 XX
 XX Claim 15; SEQ ID NO 24319; 530pp; English.
 XX
 CC The present invention relates to single exon nucleic acid probes for
 CC measuring human gene expression in a sample derived from human heart (see
 CC ABA21535-ABA411305). The present sequence is a protein encoded by one such
 CC probe. The probes may be used for predicting, measuring and displaying
 CC gene expression in samples derived from the human heart via microarrays.
 CC By measuring gene expression, the probes are useful for predicting, the
 CC diagnosing, grading, staging, monitoring and prognosing diseases of the
 CC human heart and vascular system e.g. cardiovascular disease,
 CC hypertension, cardiac arrhythmias and congenital heart disease. Note: The
 CC sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences
 XX
 XX SQ Sequence 97 AA;
 Query Match 4.8%; Score 491; DB 4; Length 97;
 Best Local Similarity 100.0%; Pred. No. 1.8e-24; Indels 0; Gaps 0;
 Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 439 SDSNESDTEGSEEDDDKQDESDDTEGEKTSMLKNTTSSVKSPMSLTGHSTPRN 498
 Db 1 SDSNESDTEGSEEDDDKQDESDDTEGEKTSMLKNTTSSVKSPMSLTGHSTPRN 60
 Qy 499 LHIAKAPGSAPAAALCSESQSPAPLGTSSSTLTSSPHS 535
 Db 61 LHIAKAPGSAPAAALCSESQSPAPLGTSSSTLTSSPHS 97
 RESULT 45
 AAM70385
 ID AAM70385 standard; protein; 97 AA.
 AC AAM70385;
 XX
 XX 06-NOV-2001 (first entry)
 DT
 DE Human bone marrow expressed probe encoded protein SEQ ID NO: 30691.
 KW Human; bone marrow expressed exon; gene expression analysis; probe;
 KW microarray; cancer; leukaemia; lymphoma; myeloma.
 XX
 OS Homo sapiens.
 XX
 XX WO200157276-A2.
 PN
 XX
 XX 09-AUG-2001.
 PD
 XX
 XX 30-JAN-2001; 2001WO-US000668.
 PF
 XX
 XX 04-FEB-2000; 2000US-0180312P.
 PR
 XX 26-MAY-2000; 2000US-0207456P.
 PR
 XX 30-JUN-2000; 2000US-00608408.
 PR
 XX 03-AUG-2000; 2000US-00632366.
 PR
 XX 21-SEP-2000; 2000US-0234687P.
 PR
 XX 27-SEP-2000; 2000US-0236359P.
 PR
 XX 04-OCT-2000; 2000GB-00024263.
 PR
 XX (MOLE-) MOLECULAR DYNAMICS INC.
 PA
 XX Penn SG, Hanzel DK, Chen W, Rank DR;
 PI WPI; 2001-48899/53.
 XX
 XX Single exon nucleic acid probes for analyzing gene expression in human
 PT brains.
 PT
 XX
 XX Example 4; SEQ ID NO 30065; 650pp + Sequence Listing; English.
 PS
 XX The present invention provides a number of single exon nucleic acid
 CC probes which are derived from genomic sequences expressed in the human
 CC brain. They can be used to measure gene expression in brain cell samples,
 CC which may enable the diagnosis and improved treatment of nervous system
 CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
 CC

CC epilepsy and cancers. The present sequence is a protein encoded by one of
 CC the probes of the invention

XX Sequence 97 AA;

Query Match 4.8%; Score 491; DB 4; Length 97;
 Best Local Similarity 100.0%; Pred. No. 1.8e-24; Mismatches 0; Indels 0; Gaps 0;
 Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 439 SDSNESDEGSEEDDDKQDESDSDTEGKTSMLNKTTSVKSPMSLTGHSTPRN 498
 Db 1 SDSNESDEGSEEDDDKQDESDSDTEGKTSMLNKTTSVKSPMSLTGHSTPRN 50
 QY 499 LHIAPGSAAPALCSESQSPAFLTSSSTLTSSPHS 535
 Db 61 LHIAPGSAAPALCSESQSPAFLTSSSTLTSSPHS 97

RESULT 47

ABG40029
 ID ABG40029 standard; peptide; 97 AA.

XX AC ABG40029;

XX DT 19-AUG-2002 (first entry)

DE Human peptide encoded by genome-derived single exon probe SEQ ID 29694.

KW Human; single exon probe; asthma; lung cancer; COPD; ILD;
 KW chronic obstructive pulmonary disease; interstitial lung disease;
 KW familial idiopathic pulmonary fibrosis; neurofibromatosis;
 KW tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;
 KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;
 KW pulmonary histiocytosis; lymphangioleiomyomatosis; Karagener syndrome;
 KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
 KW primary ciliary dyskinesia; pulmonary hypertension;
 KW hyaline membrane disease.

XX Homo sapiens.

XX WO200186003-A2.

XX PD 15-NOV-2001.

XX PF 30-JAN-2001; 2001WO-US000665.

XX PR 04-FEB-2000; 2000US-0180312P.

XX PR 26-MAY-2000; 2000US-0207456P.

XX PR 30-JUN-2000; 2000US-00608408.

XX PR 03-AUG-2000; 2000US-00632366.

XX PR 21-SEP-2000; 2000US-0234687P.

XX PR 27-SEP-2000; 2000US-0236359P.

XX PR 04-OCT-2000; 2000GB-00024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2002-114183/15.

XX Spatially-addressable set of single exon nucleic acid probes, used to

XX measure gene expression in human lung samples.

XX Claim 27; SEQ ID NO 29694; 634pp; English.

XX The invention relates to a spatially-addressable set of single exon
 CC nucleic acid probes for measuring gene expression in a sample derived
 CC from human lung comprising single exon nucleic acid probes having one of
 CC 12614 nucleic acid sequences mentioned in the specification, or their
 CC complements or the 12387 open reading frames derived from the 12614
 CC probes. Also included are a microarray comprising the novel set of probes
 CC ; the novel set of probes which hybridise at high stringency to a nucleic
 CC acid expressed in the human lung; measuring gene expression in a sample

CC derived from human lung, comprising (a) contacting the array with a
 CC collection of detectably labeled nucleic acids derived from human lung
 CC mRNA, and (b) measuring the label detectably bound to each probe of the
 CC array; identifying exons in a eukaryotic genome, comprising (a)
 CC algorithmically predicting at least one exon from genomic sequences of
 CC the eukaryote; and (b) detecting specific hybridisation of detectably
 CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,
 CC having a fragment identical to the predicted exon, the probe is included
 CC in the above mentioned microarray; assigning exons to a single gene,
 CC comprising (a) identifying exons from genomic sequence by the method
 CC above and (b) measuring the expression of each of the exons in several
 CC tissues and/or cell types using hybridisation to a single exon
 CC microarrays having a probe with the exon, where a common pattern of
 CC expression of the exons in the tissues and/or cell types indicates that
 CC the exons should be assigned to a single gene; a peptide comprising one
 CC of 12011 sequences, mentioned in the specification, or encoded by the
 CC probes/open reading frames (ORF). The probes are used for gene expression
 CC analysis, and for identifying exons in a gene, particularly using human
 CC lung derived mRNA and for the study of lung diseases such as asthma, lung
 CC cancer, chronic obstructive pulmonary disease (COPD), interstitial lung
 CC disease (ILD), familial idiopathic pulmonary fibrosis, neurofibromatosis,
 CC tuberous sclerosis, Gaucher's disease, Niemann-Pick disease, Hermansky-
 CC Pudlak syndrome, sarcoidosis, pulmonary haemosiderosis, pulmonary
 CC histiocytosis, lymphangioleiomyomatosis, pulmonary alveolar proteinosis,
 CC Karagener syndrome, fibrocystic pulmonary dysplasia, primary ciliary
 CC dyskinesia, pulmonary hypertension and hyaline membrane disease. The
 CC present sequence is a peptide/protein encoded by a single exon probe of
 CC the invention. Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic format
 CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX Sequence 97 AA;

Query Match 4.8%; Score 491; DB 5; Length 97;

Best Local Similarity 100.0%; Pred. No. 1.8e-24; Mismatches 0; Indels 0; Gaps 0;

Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 439 SDSNESDEGSEEDDDKQDESDSDTEGKTSMLNKTTSVKSPMSLTGHSTPRN 498

Db 1 SDSNESDEGSEEDDDKQDESDSDTEGKTSMLNKTTSVKSPMSLTGHSTPRN 60

QY 499 LHIAPGSAAPALCSESQSPAFLTSSSTLTSSPHS 535

Db 61 LHIAPGSAAPALCSESQSPAFLTSSSTLTSSPHS 97

RESULT 48

ABBS8706

ID ABBS8706 standard; protein; 1476 AA.

XX AC ABBS8706;

XX DT 26-MAR-2002 (first entry)

XX DE Drosophila melanogaster polypeptide SEQ ID NO 2910.

XX Drosophila; developmental biology; cell signalling; insecticide;

XX pharmaceutical.

XX Drosophila melanogaster.

XX WO200171042-A2.

XX PD 27-SEP-2001.

XX PF 23-MAR-2001; 2001WO-US0009231.

XX PR 23-MAR-2000; 2000US-0191637P.

XX PR 11-JUL-2000; 2000US-00614150.

XX PA (PEXE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.
 DR N-PSDB; ABL02809.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions.
 XX
 PS Disclosure; SEQ ID NO 2910; 21pp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (AB16176-ABL30511), expressed DNA
 CC sequences (AB101840-AB16175) and the encoded proteins (ABBS7737-
 CC ABB72072). The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 1476 AA;
 Query Match 4.7%; Score 481; DB 4; Length 1476;
 Best Local Similarity 18.9%; Pred. No. 5e-22;
 Matches 310; Conservative 192; Mismatches 447; Indels 688; Gaps 55;
 QY 665 RMRRKRPNNVNAEFNDADAKLLRLKLAQETARQAQKLRKLOKQOARV----- 719
 DB 163 RLRLPNKGP-----SAE-LDLPPRLRR-----SRMEFNLENSMFTKSNVSRVDGLLR 210
 QY 720 -AKAKQQAIAAAE-----KKQKEQIKMKQOEKIKRIQIRMEKELRAQOILEA 771
 DB 211 PKPEAYKQYVTDGPNVFNFTIFGKMPRYSPAKIKPDGK-----KOSTLNKYIVAGEATAA 266
 QY 772 KKKKKEEAANAKLAEAKRIKEKMRQOAVLLKHQERRRRQHMMLKMAEARKKAEK 831
 DB 267 KSKAKAKSDAKSLAELEKRV-----REKEAKLIE-----LEKQAEKKAQLI 309
 QY 832 ERLKQEKDEKRLNKEKLEORLEMAKELKPNEDMCLADQKPLPELPRIGLVLSG 891
 DB 310 ERVENE-----CNLLQKTDLLERTDQKVLPRYQIVTL-LPE 346
 QY 892 STFSCLMVVFLRNFGLVDNIDVPLSLVQ-EGLANIGDSMGVEQDLVPLLSAA 950
 DB 347 HLLGDAFMREFMHTYTGLLS-GIEVFQNLSFYEMTALTARETAGPLSDILLVLL-GT 404
 QY 951 VCDPLGITGYKAKTALGEHLNVGNVRDNVSEILQIFMEAHCGQTETLESIKT----- 1003
 DB 405 VFD-----LQKEEBEACAVTYLDRAAQTOEPYWSM 434
 QY 1004 -KAFOAHTPAOKASVLAFLINELACSKSVSEIDK-----NIDYNSLRDRKVV----- 1052
 DB 435 AQAASHLYAKRH--FSFKVNELPLDALTSEVLRLHLGSGAFVNEKAERWRVYRNGY 492
 QY 1053 -----EGKRLRIIHAKTGK-----RDT-----SGGIDLGE----- 1081
 DB 493 SKKDPGLLEHLSHILRIILKNHSVYQLKPKDMLLIRCLMSQIMTYSGTINLIERME 552
 QY 1082 -----QHPLTGTPTPGKRKRKGGSDYDDDDDDDDQDQDEDEDEKPKQKKTIDIC 1136
 DB 553 QTAARQDLRALVVGKRLAA-----VEINRKLKTQMH 586
 QY 1137 EDEDEG-----DOAASVELEKQIEKLSKQSQYRRKLFADSHLSRS-----VNF 1181
 DB 587 HLEVNGVEPEKREALVEKLSIAELHAQSQQOHRK-----HELQMLKLSHLSQNFVLYL 641
 QY 1182 GPDVRRRYWILPRCGGIFVEGMESGEGLEETAKEREKLAESVQIKEMFETSGDSL 1241
 DB 642 GMDCRYKRYVLESMPGIFV-----EHSPPDSL----- 668
 QY 1242 CSNTDHCQKEDLKEKDNVN-----LPLOKPGSFKLSKLLEV--AKMPPESEVMTPKPNAG 1296

DB 669 ----DTC-----LEQPIITNKSQIEIROQSALPKNRKDLRVVLLKLYGDDEKTKK----- 714
 QY 1297 ANGCTLSYQNSGKHSIGSVQSTATQSNVEKADSNLNFNTGSSGKGFYSPLPNDQLLTKL 1356
 DB 715 -----KAKHSLENKENQEHRLN-----GSAEPMDVESDSPE----- 745
 QY 1357 TEKNRQWFSLLRTPPCDDTSLTHADMSTASLVTPQSQPPSKSPSPTPAPLSSAQNPVGL 1416
 DB 746 -----APTHFELLMCS----- 756
 QY 1417 NPFALSPLOVKGVSMMGLQFCGMPTGVVTSNIPTLSVPSLGSGLGSENGNSFLTSN 1476
 DB 757 -----GDKRSCIVHD 766
 QY 1477 VASSKSPVPQNEKATSAQAAVEAVKVPDFPSPKPIPEMQFGWMIILIPEDLKALK 1536
 DB 767 SRNGQRQ-----WAYIYKAEEDIDELIK 789
 QY 1537 VLHLRGIRE-----KALQKQIOKH-----LDYITQACLKNKDVAVIIELNENE----- 1579
 DB 790 ALNPNGLREYELLQELSVLRSLEIQQHAKTCVPDILLSLENETMRKKFMAAMESETNRKYGE 849
 QY 1580 -----NQVTR-DIVEN-WSVEQAMEMDLVLQOVEDLER-----RVASA 1617
 DB 850 ANFGLNGTDLNEVRLHLVDRIIOFENDIYTGDLGRL-KVKDMEKMRSDLLGGNYDAQ 908
 QY 1618 SLQVKGWCMPEPASEREDLVYFEHKSFTKLCKEHDGEFTG----- 1657
 DB 909 KIQ-----W-----GPGKLEDEAGSDNESH-ETHEEDDGALLGKYARKPYRDPGMYLAASADT 961
 QY 1658 -----EDSSAHALERKSDNPLDIIV-----TRLADLERNIERIEED----- 1695
 DB 962 KPLPDSDEEHOHT-----NAVSIPTAVHNASALLOVEQAIGKRFKEPYGMKKNDPK 1015
 QY 1696 -----IAPGLRVRRRALSARSAAQVALCIIQOLQKSIAMSEKSIIMKYVCOICRKGDNEE 1748
 DB 1016 QEALKLACDSRLHQWESVLMESTSFAQVFLHLNHLHDCIQWRRSTNKSCLKVCRGSDPE 1075
 QY 1749 LLLLCDGCDKCHYCHRPKITTPDGWDFCPACIAK---ASGQTLKIKLHVKGKK--- 1802
 DB 1076 KMLLCDCECNAGTHMFLCKPLKRSVPPGPNWYCNDCVKSLGSLNGONEKDKQATKKRKF 1135
 QY 1803 -----TNESKGGK-VTLTG-----DTEDEDSASTSSSLKRGN----- 1834
 DB 1136 VEEDDEATDEEBEKKDDMTDEDAEHENKHDVEDDESVTSTPSSRVNGRILRRP 1195
 QY 1835 -----KDLQKRMKEENTSINLSKQESFTSVKPKPRDDSKDLALCSM----- 1875
 DB 1196 RTRPTSRRLTSKEIEEHAQEDVDGSDVDDASLTAGEDTIEDSEDEKVCQKCFYDGEI 1255
 QY 1876 ----- 1875
 DB 1256 KVCQRLFFHLECVHLKRPRTDFVCKTKPMQPRRRRHNSNMNGDHRDDEEPKAKRPR 1315
 QY 1876 -----ILTEMETH 1883
 DB 1316 NSLRISLDTKARPNGNN 1375
 QY 1884 EDAMPFLPVLNKLVPYKVKIKKPMDFSTIREKLSGQYPNLETTFADVLVFNPCBTF 1943
 DB 1376 KAAWFLRPVLTSVPDYHQIITKTPMDLAKIKSKLNMGAYQLNEELLSDIQLVFRNCOLY 1435
 QY 1944 NEDSDIGRAGHNRKY 1960
 DB 1436 NVEGNEIYDAGCOLERF 1452
 RESULT 49
 AAM17059
 ID AAM17059 standard; protein; 91 AA.
 XX
 AC AAM17059;
 XX

```

DT 12-OCT-2001 (first entry)
XX Peptide #3493 encoded by probe for measuring cervical gene expression.
DE Probe; human; microarray; gene expression; cervical epithelial cell;
XX cervical cancer.
KW Homo sapiens.
XX WO200157278-A2.
XX 09-AUG-2001.
XX 30-JAN-2001; 2001WO-US000670.
XX 04-FEB-2000; 2000US-0180312P.
XX 26-MAY-2000; 2000US-0207456P.
XX 30-JUN-2000; 2000US-00608408.
XX 03-AUG-2000; 2000US-00632366.
XX 21-SEP-2000; 2000US-0234687P.
XX 27-SEP-2000; 2000US-0236359P.
XX 04-OCT-2000; 2000GB-00024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488901/53.
XX Human genome-derived single exon nucleic acid probes, useful for analyzing
XX gene expression in human cervical epithelial cells.
XX Claim 27; SEQ ID NO 21885; 487pp; English.
XX The present invention relates to human single exon nucleic acid probes
XX (SENP: see A110068-A128459). The present sequence is a peptide encoded
XX by one such probe. The SENPs are derived from human HeLa cells. The SENPs
XX can be used to produce a single exon microarray, which can be used for
XX measuring human gene expression in a sample derived from human cervical
XX epithelial cells. By measuring gene expression, the probes are therefore
XX useful in grading and/or staging of diseases of the cervix, notably
XX cervical cancer. Note: The sequence data for this patent did not form
XX part of the printed specification, but was obtained in electronic format
XX directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 91 AA;
XX
XX Query Match 4.6%; Score 463; DB 4; Length 91;
XX Best Local Similarity 98.9%; Pred. No. 1.1e-22;
XX Matches 90; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
XX 1426 VKGGVSMGLQFCGWTGVTNSIPFTLSVPSLGLSGGLSGGNSFLTSNVASSKSESP 1485
XX 1 VKGGVSMGLQFCGWTGVTNSIPFTLSVPSLGLSGGLSGGNSFLTSNVASSKSESP 60
XX
XX 1486 VPQNEKATSAQPAAVEAKPVDPPSPKPIPE 1516
XX 61 VPQNEKATSAQPAAVEAKPVDPPSPKPIPE 91
XX
XX RESULT 50
XX ABB36058
XX ID ABB36058 standard; peptide; 91 AA.
XX AC ABB36058;
XX XX
XX 04-FEB-2002 (first entry)
XX Peptide #3564 encoded by human foetal liver single exon probe.
XX Human; foetal liver; gene expression; single exon nucleic acid probe.
XX Homo sapiens.

```

```

XX WO200157277-A2.
XX 09-AUG-2001.
XX 30-JAN-2001; 2001WO-US000669.
XX 04-FEB-2000; 2000US-0180312P.
XX 26-MAY-2000; 2000US-0207456P.
XX 30-JUN-2000; 2000US-00608408.
XX 03-AUG-2000; 2000US-00632366.
XX 21-SEP-2000; 2000US-0234687P.
XX 27-SEP-2000; 2000US-0236359P.
XX 04-OCT-2000; 2000GB-00024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-483447/52.
XX Human genome-derived single exon nucleic acid probes useful for analyzing
XX gene expression in human fetal liver.
XX Claim 27; SEQ ID NO 28693; 639pp + Sequence Listing; English.
XX The invention relates to a single exon nucleic acid probe for measuring
XX human gene expression in a sample derived from human foetal liver. The
XX single exon nucleic acid probes may be used for predicting, measuring and
XX displaying gene expression in samples derived from human fetal liver. The
XX present sequence is a peptide encoded by a single exon nucleic acid probe
XX of the invention. Note: The sequence data for this patent did not form
XX part of the printed specification, but was obtained in electronic format
XX directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 91 AA;
XX
XX Query Match 4.6%; Score 463; DB 4; Length 91;
XX Best Local Similarity 98.9%; Pred. No. 1.1e-22;
XX Matches 90; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
XX 1426 VKGGVSMGLQFCGWTGVTNSIPFTLSVPSLGLSGGLSGGNSFLTSNVASSKSESP 1485
XX 1 VKGGVSMGLQFCGWTGVTNSIPFTLSVPSLGLSGGLSGGNSFLTSNVASSKSESP 60
XX
XX 1486 VPQNEKATSAQPAAVEAKPVDPPSPKPIPE 1516
XX 61 VPQNEKATSAQPAAVEAKPVDPPSPKPIPE 91
XX
XX Search completed: March 9, 2005, 14:05:50
XX Job time : 173.925 secs

```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 9, 2005, 13:18:56 ; Search time 158.583 Seconds
(without alignments)
5405.490 Million cell updates/sec

Title: US-10-702-148-1
Perfect score: 8631
Sequence: 1 MEDASRSGVAPLNNVL.....VTPSNVDQVTPPAKXSR1 1674

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 65 summaries

Database : UniProt_03.*

1: uniprot_eprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7789	90.2	1556	1	BA1A_HUMAN
2	3793	43.9	892	2	Q6P1D9
3	3098.5	35.9	718	2	Q8C8D1
4	1691	19.6	627	1	BA1A_XENLA
5	1501	17.4	344	1	BA1A_MOUSE
6	1424.5	16.5	1357	2	Q7PRP9
7	1280.5	14.8	1476	2	Q9Y9T4
8	1279.5	14.8	1476	2	Q9Y0W1
9	1275.5	14.8	1476	2	Q9NG24
10	1228	14.2	1202	2	Q7QBK2
11	1194	13.8	371	2	Q66IE0
12	1176	13.6	1427	2	Q9N5L9
13	945.5	11.0	1479	1	BA1B_MOUSE
14	941.5	10.9	1483	1	BA1B_HUMAN
15	838	9.7	1079	2	Q8UVR4
16	756.5	8.8	2130	1	BA2B_CHICK
17	731.5	8.5	1972	1	BA2B_HUMAN
18	695.5	8.1	1878	1	BA2A_HUMAN
19	689.5	8.0	1905	2	Q68DI8
20	673	7.8	1698	2	Q6VI94
21	657.5	7.6	1850	1	BA2A_MOUSE
22	614.5	7.1	2060	2	Q7Q3S9
23	591.5	6.9	3109	2	Q9BMQ0
24	591	6.8	3080	2	Q9V602
25	576.5	6.7	657	2	Q8CAU9
26	530	6.1	1376	2	Q23590
27	482	5.6	723	2	Q9FNM6
28	474	5.5	955	2	Q86UJ6
29	460.5	5.3	886	2	Q80U42
30	441	5.1	617	2	Q6Z178
31	408	4.7	1230	2	Q6BV90

ALIGNMENTS

RESULT 1

ID	BA1A_HUMAN	STANDARD;	PRT; 1556 AA.
AC	Q9NRL2; Q9NZ15; Q9P065; Q9UIG1; Q9Y3V3;		
DT	28-FEB-2003 (Rel. 41, Created)		
DT	28-FEB-2003 (Rel. 41, Last sequence update)		
DT	05-JUL-2004 (Rel. 44, Last annotation update)		
DE	Bromodomain adjacent to zinc finger domain protein 1A (ATP-utilizing		
DE	chromatin assembly and remodeling factor 1) (hACF1) (ATP-dependent		
DE	chromatin remodeling protein) (Williams syndrome transcription		
DE	factor-related chromatin remodeling factor 180) (WCRF180) (hWALp1)		
DE	(CHRAC subunit ACF1) (HSPC317).		
GN	Name=BAZ1A; Synonyms=ACF1, WCRF180;		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A. (ISOFORM 1), AND MASS SPECTROMETRY.		
RC	TISUS=Cervical carcinoma;		
RA	MEDLINE=20341093; PubMed=1080450; DOI=10.1093/emboj/19.13.3377;		
RA	Post R.A., Dallaire G., Huelsmann B.B., Grimaldi M.A., Corona D.F.V.,		
RA	Becker P.B., Bickmore W.A., Varga-Weisz P.D.;		
RT	"HuCHAC, a human ISWI chromatin remodeling complex contains hACF1		
RT	and two novel histone-fold proteins.";		
RL	EMBO J. 19:3377-3387(2000).		
RN	[2]		
RP	SEQUENCE FROM N.A. (ISOFORM 2).		
RX	MEDLINE=20122567; PubMed=10655480; DOI=10.1073/pnas.97.3.1038;		
RA	Bochar D.A., Savard J., Wang W., Lafleur D.W., Moore P., Cote J.,		
RA	Shiekhattar R.;		
RT	"A family of chromatin remodeling factors related to Williams syndrome		
RT	transcription factor.";		
RL	Proc. Natl. Acad. Sci. U.S.A. 97:1038-1043(2000).		

[3]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE-Testis;
RX MEDLINE=20130112; PubMed=10662543; DOI=10.1006/geno.1999.6071;
RA Jones M.H., Hamana N., Nezu J.-I., Shimane M.;
RT "A novel family of bromodomain genes.";
RL Genomics 63:40-45(2000).
[4]
RP SEQUENCE OF 1-541 FROM N.A. (ISOFORM 1).
RC TISSUE=Blood;
RA Ye M., Zhang Q.H., Zhou J., Shen Y., Wu X.Y., Guan Z.Q., Wang L.,
RA Fan H.Y., Mao Y.F., Dai M., Huang Q.H., Chen S.J., Chen Z.;
RT "Human partial CDS from cd34+ stem cells.";
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
[5]
RP SEQUENCE OF 746-1556 FROM N.A. (ISOFORM 1).
RC TISSUE=Uterus;
RA Wambutt R., Heubner D., Mewes H.-W., Gassenhuber J., Wiemann S.;
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: May play a role in transcriptional regulation. May be
involved in the formation or maintenance of heterochromatin
playing a critical role in developmental control.
CC -!- SUBUNIT: Together with CHRA1, POE3 and ISWI/SNF2H proteins, it
forms the ISWI chromatin remodelling complex, CHRA1.
CC -!- SUBCELLULAR LOCATION: Nuclear. May target the CHRA1 complex to
heterochromatin.
CC -!- ALTERNATIVE PRODUCTS:
CC Event-Alternative splicing; Named isoforms=2;
CC Name=1;
CC IsoId=Q9NRL2-1; Sequence=Displayed;
CC Name=2;
CC IsoId=Q9NRL2-2; Sequence=VSP_000551;
CC -!- TISSUE SPECIFICITY: Highly expressed in testis and at low or
undetectable levels in other tissues analyzed.
CC -!- MISCELLANEOUS: Stimulated by double-stranded DNA and nucleosomal
DNA.
CC -!- SIMILARITY: Belongs to the WAF family.
CC -!- SIMILARITY: Contains 1 bromodomain.
CC -!- SIMILARITY: Contains 1 DDT domain.
CC -!- SIMILARITY: Contains 1 PHD-type zinc finger.
CC -!- SIMILARITY: Contains 1 WAC domain.
CC -!- CAUTION: Ref.4 sequence differs from that shown due to multiple
frameshifts starting at position 382.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>
or send an email to license@ebi.ac.uk).
CC -----
DR EMBL; AF213467; AAF70601.1; -;
DR EMBL; AF221130; AAF32366.1; -;
DR EMBL; AB032252; BAA89209.1; -;
DR EMBL; AF161435; AAF28995.1; ALT_SEQ.
DR EMBL; AL050089; CAB43261.1; -;
DR PIR; T08738; T08738.
DR HSSP; Q9UIG0; 1F62.
DR Genew; HGNC:960; BAZ1A.
DR MIM; 605680; -;
DR InterPro; IPR001487; Bromodomain.
DR InterPro; IPR004022; DDT.
DR InterPro; IPR001965; Znf PHD.
DR InterPro; IPR001841; Znf ring.
DR Pfam; PF00439; Bromodomain; 1.
DR Pfam; PF02791; DDT; 1.
DR Pfam; PF00628; PHD; 1.
DR PRINTS; PR00503; BROMODOMAIN.
DR SMART; SM00297; BROMO; 1.
DR SMART; SM00571; DDT; 1.
DR SMART; SM00249; PHD; 1.
DR SMART; SM00184; RING; 1.

DR PROSITE; PS00633; BROMODOMAIN_1; 1.
DR PROSITE; PS50014; BROMODOMAIN_2; 1.
DR PROSITE; PS50827; DDT; 1.
DR PROSITE; PS01359; ZF_PHD_1; 1.
DR PROSITE; PS50016; ZF_PHD_2; 1.
KW Alternative splicing; Bromodomain; Coiled coil; Nuclear protein;
KW Transcription regulation; Zinc-finger.
FT DOMAIN 22 128 WAC.
FT DOMAIN 422 487 DDT.
FT ZN_FING 1148 1198 PHD-type.
FT DOMAIN 1446 1516 Bromodomain.
FT DOMAIN 306 397 Coiled coil (Potential).
FT DOMAIN 634 709 Coiled coil (Potential).
FT VARSPIC 504 535 Missing (in isoform 2).
FT /FTid=VSP_000551.
FT CONFLICT 135 135 R -> T (in Ref. 3).
FT CONFLICT 236 236 D -> E (in Ref. 3).
FT CONFLICT 551 551 D -> E (in Ref. 2 and 3).
FT CONFLICT 730 730 V -> F (in Ref. 3).
FT CONFLICT 769 769 P -> L (in Ref. 3).
FT CONFLICT 1201 1201 S -> C (in Ref. 3).
FT CONFLICT 1206 1206 S -> F (in Ref. 3).
FT CONFLICT 1366 1366 K -> N (in Ref. 2 and 3).
FT CONFLICT 1409 1409 K -> R (in Ref. 2 and 3).
SQ SEQUENCE 1556 AA; 178674 MW; 43P2667C3CCFD458 CRC64;

Query Match 90.2%; Score 7789; DB 1; Length 1556;
Best Local Similarity 97.4%; Pred. No. 1.2e-312;
Matches 1515; Conservative 3; Mismatches 6; Indels 32; Gaps 1;

Qy 151 MPLLHRRPFVRQPPADLRPDEEVFYCKVTNEIFRHYDDFFERTILCSLWSCAVTGRP 210
Db 1 MPLLHRRPFVRQPPADLRPDEEVFYCKVTNEIFRHYDDFFERTILCSLWSCAVTGRP 60

Qy 211 GLTYQEALESEKKARQNLQSPFPEPLIIPVLYLSLTHRSRLHEICDDIFAYVKDVFVEE 270
Db 61 GLTYQEALESEKKARQNLQSPFPEPLIIPVLYLSLTHRSRLHEICDDIFAYVKDVFVEE 120

Qy 271 TVEVIRNGARLOCTILEVLPPSHONGFANGHNSVDGETIIISDSSETQSCSFONGK 330
Db 121 TVEVIRNGARLOCTILEVLPPSHONGFANGHNSVDGETIIISDSSETQSCSFONGK 180

Qy 331 KKDAIDPLLFYKVPQTKKELHESAIKATQISRRKHLFSRDKLFLKQHCPEQGVIK 390
Db 181 KKDAIDPLLFYKVPQTKKELHESAIKATQISRRKHLFSRDKLFLKQHCPEQGVIK 240

Qy 391 IKASSLSTYKIAEODFSYFFPDPPPTFIFSPANRRGRPPKRIHISOEDNVANKQTILASY 450
Db 241 IKASSLSTYKIAEODFSYFFPDPPPTFIFSPANRRGRPPKRIHISOEDNVANKQTILASY 300

Qy 451 RSKATKERDILLKQEEEMKSLAFKAKLREKADALEAKKEKEDKKEKRELKIVVEER 510
Db 301 RSKATKERDILLKQEEEMKSLAFKAKLREKADALEAKKEKEDKKEKRELKIVVEER 360

Qy 511 LKKKEERLKVREKEREKREKRYVEYLQWSPREDMECDLLELPEPTPVKTRL 570
Db 361 LKKKEERLKVREKEREKREKRYVEYLQWSPREDMECDLLELPEPTPVKTRL 420

Qy 571 PPEIFGDALMVLEFLNAGFELFDLQDEFDPDGVTLVELEALVGNDSGGPCELLFFFLTA 630
Db 421 PPEIFGDALMVLEFLNAGFELFDLQDEFDPDGVTLVELEALVGNDSGGPCELLFFFLTA 480

Qy 631 IFQIAEEEEVEAKEQLTDADTK-----GCSLK 658
Db 481 IFQIAEEEEVEAKEQLTDADTK-----GCSLK 540

Qy 659 SLDLDSCTLSLILRLHLTLAGADVTSANAKYRQKGGFDATDDACMELRLSNPLSVKVL 718
Db 541 SLDLDSCTLSLILRLHLTLAGADVTSANAKYRQKGGFDATDDACMELRLSNPLSVKVL 600

Qy 719 SSTSVYDLTPCEKMKILHALCGKLLTLVSTRDFIEDYVDILRQAKQSFRELKAEQHRKER 778
Db 601 SSTSVYDLTPCEKMKILHALCGKLLTLVSTRDFIEDYVDILRQAKQSFRELKAEQHRKER 660

QY	779	EEAAARIRKKEEKLKEQEQMKKEQKLKEDEQNSTADISIGEEREDFDTTIESKOT	838
Db	661	EEAAARIRKKEEKLKEQEQMKKEQKLKEDEQNSTADISIGEEREDFDTTIESKOT	720
QY	839	EQKELDDMFTEDEDDPGSHKRGKRGKQNGKFEFTQEQINCVTRRELLTADDEEALKQ	898
Db	721	EQKELDDMFTEDEDDPGSHKRGKRGKQNGKFEFTQEQINCVTRRELLTADDEEALKQ	780
QY	899	EHQKKEKELLEKIQSAIACTNIFPLGRDRMYRRYWFPSIFGLFTEEDYSGLTEDMLLPR	958
Db	781	EHQKKEKELLEKIQSAIACTNIFPLGRDRMYRRYWFPSIFGLFTEEDYSGLTEDMLLPR	840
QY	959	PSSFQNNVQSDPQVSTTKTGBPLMSSTNIDQGRDRHSVOLPKPVKPNRWCFYSSCEQ	1018
Db	841	PSSFQNNVQSDPQVSTTKTGBPLMSSTNIDQGRDRHSVOLPKPVKPNRWCFYSSCEQ	900
QY	1019	LDQILEALNSRGHRESALKETLLOEKSRICQAQARFSEKHFSDKPODPSKPTYSRGRS	1078
Db	901	LDQILEALNSRGHRESALKETLLOEKSRICQAQARFSEKHFSDKPODPSKPTYSRGRS	960
QY	1079	SNAYDPSQWCAEQLELRLDFLDIEDRIYQGTILGAIKVTDRHIWRSALSGRYELLSE	1138
Db	961	SNAYDPSQWCAEQLELRLDFLDIEDRIYQGTILGAIKVTDRHIWRSALSGRYELLSE	1020
QY	1139	ENKNGIITKTVNEDVEEIMIDEQTKVIVKDRLLGLTKTPTSTVSTNASTPOSVSVVHYL	1198
Db	1021	ENKNGIITKTVNEDVEEIMIDEQTKVIVKDRLLGLTKTPTSTVSTNASTPOSVSVVHYL	1080
QY	1199	AWALFOIQEITERRFLKAPLADSDSGRSYKTVLDRWRESLLSSASLSQVFLHSLTDRSV	1258
Db	1081	AWALFOIQEITERRFLKAPLADSDSGRSYKTVLDRWRESLLSSASLSQVFLHSLTDRSV	1140
QY	1259	IWSKILNARCKICRKGDAENMWLDCDCDRGHHTYCVRPKLTVPEDGDFCPCRPKOR	1318
Db	1141	IWSKILNARCKICRKGDAENMWLDCDCDRGHHTYCVRPKLTVPEDGDFCPCRPKOR	1200
QY	1319	CRRLSFRQPSLEDEVEDSMGDEDDVDGDEEGQSEEEYEVEQDEDDDSQEEEEVSL	1378
Db	1201	SRRLSRRQPSLEDEVEDSMGDEDDVDGDEEGQSEEEYEVEQDEDDDSQEEEEVSL	1260
QY	1379	PKRGEPQVRLPVKTRGKLSFSSFRGQOQEGRPSPRSQQSTPKTTVSSKTGRSLRKINS	1438
Db	1261	PKRGEPQVRLPVKTRGKLSFSSFRGQOQEGRPSPRSQQSTPKTTVSSKTGRSLRKINS	1320
QY	1439	APPTETKSLRIASRSTRHSHGFLQADVVELLSPRKRRGRKSANNTPENSPNPNFVI	1498
Db	1321	APPTETKSLRIASRSTRHSHGFLQADVVELLSPRKRRGRKSANNTPENSPNPNFVI	1380
QY	1499	ATKSSEQSRVNIASKLSQESKRRCKRQSPSPVTLGRSSRGQGVHLSAFEQ	1558
Db	1381	ATKSSEQSRVNIASKLSQESKRRCKRQSPSPVTLGRSSRGQGVHLSAFEQ	1440
QY	1559	LVLVLVRHDDSWPFLKLVSKIQVDPYDIIKKPALTALNIIRKVNKCEYKLASEFIDDIEL	1618
Db	1441	LVLVLVRHDDSWPFLKLVSKIQVDPYDIIKKPALTALNIIRKVNKCEYKLASEFIDDIEL	1500
QY	1619	MFSNCFEYNPNRTSEAKAGTFLQAFHHIQAOKLGLHVTSPNVDVSTPPAAKSKRI	1674
Db	1501	MFSNCFEYNPNRTSEAKAGTFLQAFHHIQAOKLGLHVTSPNVDVSTPPAAKSKRI	1556
RESULT 2			
Q6PID9			
ID	Q6PID9	PRELIMINARY;	PRT; 892 AA.
AC	Q6PID9;		
DT	05-JUL-2004 (TReMBLrel. 27, Created)		
DT	05-JUL-2004 (TReMBLrel. 27, Last sequence update)		
DE	05-JUL-2004 (TReMBLrel. 27, Last annotation update)		
DE	BC065123 protein (Fragment).		
GN	Name=BC065123;		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		

OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_TaxID=10090;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=C57BL/6; TISSUE=Brain;		
RX	MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;		
RA	Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,		
RA	Klausner R.D., Collins P.S., Wagner L., Shennan C.M., Schuler G.D.,		
RA	Altshul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,		
RA	Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,		
RA	Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,		
RA	Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,		
RA	Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,		
RA	Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,		
RA	Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,		
RA	Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,		
RA	Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,		
RA	Fahy J., Helton E., Kettaman M., Madan A., Rodrigues S., Sanchez A.,		
RA	Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,		
RA	Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,		
RA	Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,		
RA	Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,		
RA	Jones S.J., Marra M.A.		
RT	"Generation and initial analysis of more than 15,000 full-length human		
RT	and mouse cDNA sequences";		
RL	Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=C57BL/6; TISSUE=Brain;		
RA	Strausberg R.;		
RL	Submitted (JAN-2004) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; BC065123; AAH65123.1; -.		
DR	GO; GO:0003677; F:DNA binding; IEA.		
DR	GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.		
DR	InterPro; IPR001487; Bromodomain.		
DR	InterPro; IPR011011; FIVE_PHD_Znf.		
DR	InterPro; IPR001965; Znf_PHD.		
DR	Pfam; PF00439; Bromodomain; 1.		
DR	Pfam; PF00628; PHD; 1.		
DR	PRINTS; PR00503; BROMODOMAIN.		
DR	SMART; SM00297; BROMO; 1.		
DR	SMART; SM00249; PHD; 1.		
DR	PROSITE; PS00633; BROMODOMAIN_1; 1.		
DR	PROSITE; PS0014; BROMODOMAIN_2; 1.		
DR	PROSITE; PS01359; 2F_PHD_1; 1.		
DR	PROSITE; PS0016; 2F_PHD_2; 1.		
FT	NON_TER 1		
SQ	SEQUENCE 892 AA; 101830 MW; 95A6C798F37AA93F CRC64;		
Query Match 43.9%; Score 3793; DB 2; Length 892;			
Best Local Similarity 82.5%; Pred. No. 2.3e-148;			
Matches 740; Conservative 62; Mismatches 89; Indels 6; Gaps 5;			
Qy	779	EEAAARIRKKEEKLKEQEQMKKEQKLKEDQNSTADISIGEEREDFDTTIESKOT	838
Db	1	EATAARIRRRKEEKLKEQEQMKKEQKLKEDQNSAA--VPGEEEREDFDTSTENKI	58
Qy	839	EQKELDDMFTEDEDDPGSHKRRGRKRGKNGKFEFTQEQIN-CVTRRELLTADDEEALK	897
Db	59	EQKDLDPDVVTEDEDDPGSHKRRGRKVGQAVKQCIQKEEMNYCIQPLPSADAEALR	118
Qy	898	QEHQKKEKELLEKIQSAIACTNIFPLGRDRMYRRYWFPSIFGLFTEEDYSGLTEDMLLP	957
Db	119	QEQQQKEKELDKIQSAIACTNIFPLGRDRLYRRYWFPSIFGLFTEEDYSGLTEDMLLP	178
Qy	958	RPSFQNNVQSDPQVSTTKTGBPLMSSTNIDQGRDRHSVQLPKPVKPNRWCFYSSCE	1017
Db	179	RPSFHNNAPRDPQVSIKTEESPLSESTSLDQGPFDSDVLLPKPVKPNRWCFYSSCA	238
Qy	1018	QLDQILEALNSRGHRESALKETLLOEKSRICQAQARFSEKHFSDKPODPSKPTYSGR	1077
Db	239	QLDQLIDALNSRGHRESALKETLLOEKSRICQAQLAHFSEKHFSDKPOADSKPVSSRGR	298

Qy 1078 SSNAYPSCMAEQLELRDLFLDIEDRIYQGTGLGAIKVTDRIHWSALESGRYELLS 1137
Dy 299 SSGACDISQMSAERQLELRDLFLDIEDRIYQGTGLGAIKVTDRIHWSALESGRYELLS 358
Qy 1138 BENKENGIIKTVNEDVEMETDEQTKVTKDRLLGKITETPTSTVSTNASTPQSVSVVHY 1197
Dy 359 EESKENGVIKTVNEDVEMEM-EGARVTVRDLGLKITETPTSTSTASTPQSVSVVHY 417
Qy 1198 LAMALFOIEQIERRELKAPLDADSGRSYKTVLDRWRSELSSASISQVFLHJLSTLDRS 1257
Dy 418 LALALFOIEQIERRELKAPLDGNDGSGRSYKTVLDRWRSELSSASISQVFLHJLSTLDRS 477
Qy 1258 VIWKSILNARCKICRKGDAENMVLCDGCRGHHTYCVRPKLTVPDGFCECPKQ 1317
Dy 478 VWKSILNARCKICRKGDAENMVLCDGCRGHHTYCVRPKLTVPDGFCECPKQ 537
Qy 1318 RCRLSFRQPSLSDDEVDSDMGEDDDEVGDSEEGSSEEEYEVQDEDDSDQEEBVS 1377
Dy 538 RSRLSFRQPSLSDDEMEGEMDDDEVDDDEEGSSEEEYEVQDEDDSDQEEBVS 597
Qy 1378 LPKGRPOVRLPVKTRGKLSFSFSSRGQOQEPGRYPBSRSQOSTPKTTVSSKTGSLRKIN 1437
Dy 598 PPKGRPOVRLPVKTRGKLSFSFSSRGQOQEPGRYPBSRSQOSTPKNTAKS-ASKNLKTR 656
Qy 1438 SAPPTETKSLRIASRSTRHSHGCLQADVVELLSPRKRGRKGSANNTPENSPNFPNFRV 1497
Dy 657 SAPPTETSLVGRSTHSPSALQ-DVVELLSPHSKGRGKGDHTPHSPSFTNFRV 715
Qy 1498 IATKSSQSRVNTASKLSQESSKRCRCRKRQSPSPVTLGRRSGRQGVHLSAPE 1557
Dy 716 STSRSSROLPLNTAESLSLQSSSKRCRCRKRQSTESSPVLNRRSSGRQGVHLSAPE 775
Qy 1558 QLVVELVRHDSWPKLVSKIOVPDYDIIKIPALNIREKNKCEYKLASEFIDIE 1617
Dy 776 QLVVELVRHDSWPKLVSKIOVPDYDIIKIPALNIREKNKCEYKLASEFIDIE 835
Qy 1618 LMFNCFEYNPRNTSEAKAGTRLOAFHIOAQKLGHLVTPSNVDVSTPPAAKKSRI 1674
Dy 836 LMFNCFEYNPRNTSEAKAGTRLOAFHIOAQKLGHLVSPVTDVSTPPAAKKSRI 892
RESULT 3
ID Q8C8D1 PRELIMINARY; PRT; 718 AA.
AC Q8C8D1,
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Mus musculus 10 days neonate cerebellum cDNA, RIKEN full-length
DE enriched library, clone:B910060C03 product:hypothetical Bromodomain/
DE PHD-finger containing protein, full insert sequence.
GN Name=BC065123;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_taxid=10090;
RN [1]
RP
RC STRAIN=C57BL/6J; TISSUE=Cerebellum;
RC MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RC Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning.";
RL Meth. Enzymol. 303:19-44(1999).
RN [2]
RP
RC STRAIN=C57BL/6J; TISSUE=Cerebellum;
RC MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RC RIKEN FANTOM Consortium;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
RN [3]
RP
RC STRAIN=C57BL/6J; TISSUE=Cerebellum;

RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
RN [4]
RP
RC STRAIN=C57BL/6J; TISSUE=Cerebellum;
RC MEDLINE=2049374; PubMed=11042159; DOI=10.1101/gr.145100;
RC Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA Konno H., Akizawa J., Nishi K., Kitsuai T., Tashiro H., Itoh M.,
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
RT sequencing pipeline with 384 multicapillary sequencer.";
RL Genome Res. 10:175-177(2000).
RN [6]
RP
RC STRAIN=C57BL/6J; TISSUE=Cerebellum;
RC MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RC Shibata K., Itoh M., Aizawa K., Nagao S., Sasaki N., Carninci P.,
RA Konno H., Akizawa J., Nishi K., Kitsuai T., Tashiro H., Itoh M.,
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
RT sequencing pipeline with 384 multicapillary sequencer.";
RL Genome Res. 10:175-177(2000).
RN [6]
RP
RC STRAIN=C57BL/6J; TISSUE=Cerebellum;
RC MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RC Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
RA Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,
RA Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
RA Katoh H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,
RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
RA Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,
RA Saito K., Saichou H., Sakai C., Sakai K., Sakazume N., Sano H.,
RA Sasaki D., Shibata K., Shingawa A., Shiraki T., Sogabe Y., Tagami M.,
RA Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,
RA Tomaru A., Toyota T., Yasunishi A., Muramatsu M., Hayashizaki Y.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK047427; BAC33055.1; -.
DR HSSP; Q9UIG0; I562.
DR MGD; MGI:3039627; BC065123.
DR GO; GO:0000151; C:ubiquitin ligase complex; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0004842; F:ubiquitin-protein ligase activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0016567; P:protein ubiquitination; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR001487; Bromodomain.
DR InterPro; IPR011011; FYVE PHD_ZnF.
DR InterPro; IPR001965; ZnF_PHD.
DR InterPro; IPR001841; ZnF_ring.
DR Pfam; PF00439; Bromodomain; 1.
DR Pfam; PF00628; PHD; 1.
DR PRINTS; PR00503; BROMODOMAIN.
DR SMART; SM00297; BROMO; 1.
DR SMART; SM00249; PHD; 1.
DR SMART; SM00184; RING; 1.
DR PROSITE; PS00633; BROMODOMAIN_1; 1.
DR PROSITE; PS00014; BROMODOMAIN_2; 1.
DR PROSITE; PS01359; ZF_PHD_1; 1.
DR PROSITE; PS00016; ZF_PHD_2; 1.
KW Hypothetical protein.
SQ
SEQUENCE 718 AA; B508483BEB24FE02 CRC64;
Query Match 35.9%; Score 3098.5; DB 2; Length 718;
Best local similarity 83.6%; Pred. No. 7.3e-120;
Matches 603; Conservative 46; Mismatches 69; Indels 3; Gaps 3;

```

QY 954 MLLPRPSSFQNVQSDPOVTKTGEPLMSESTSNIDQPRHSHVQLPKPVHKPNRWCFY 1013
Db 1 MLLPRPSSFQNVQSDPOVTKTGEPLMSESTSNIDQPRHSHVQLPKPVHKPNRWCFY 60
QY 1014 SSCQQLDQILALNSRGHRESALKETLLOEKSRICAQLARFSEKPFSDKPPQDSKPTY 1073
Db 61 SSCQQLDQILALNSRGHRESALKETLLOEKSRICAQLARFSEKPFSDKPPQDSKPTY 120
QY 1074 SGRGSSNAVDPQSCAQLKLELRLRDLFLDIEDRIYQGTGLGAIKVTDRHWRSALESGRY 1133
Db 121 SGRGSSGACDISQMSAERQLELRLRDLFLDIEDRIYQGTGLGAIKVTDRHWRSALENGRY 180
QY 1134 ELLSEENKENGIIKTWNEDVEEMEDQTKVTKVDRLLGIKTETSTVSTNASTQSVSS 1193
Db 181 ELLSEENKENGIIKTWNEDVEEMEM-EQARVTVDRLLGIKTETSTTSTASTPQSVSN 239
QY 1194 VVHYLAMALFOIEQGIERRFLKAPLDASDSGRSYKTVLDRWRRESLLSASLSOVFLHLST 1253
Db 240 VVHYLAMALFOIEQGIERRFLKAPLDGNDGSRSYKTVLDRWRRESLLSASLSOVFLHLST 299
QY 1254 LDRSVIWSKSIILNARCKICRKGDAENMVLCDCDRGHHTYCVRPKLKTVPEGDWFCPEC 1313
Db 300 LDRSVIWSKSIILNARCKICRKGDAENMVLCDCDRGHHTYCVRPKLKAVPDGDWFCPEC 359
QY 1314 RPKQCRRLSPRPSLESDEVEDSMGCGEDVDGDBEGQSEEEYEVQDEDDSQEE 1373
Db 360 RPKQSRRLSGRPSLESDEEMEGMEDDDDDVEDDDDEEGQSEEEYEVQDEDDSDDD 419
QY 1374 EEVSUPKGRPOVRLPVKTRGKLSSFSRSGQQEPGRYPSPSQSTPTKTVSSKTGRSL 1433
Db 420 EALSPPKGRQVRLPIKTRGFGFSRSGRQDPGRYPSPSQSTPTAKS-ASKNL 478
QY 1434 RKINSAPPTETKSLRIASRSTRSHSGPLQADVVELLSPRKRGKRSANNTPSPNFP 1493
Db 479 RKTRSAPPTETSLAVGSRSTRSHSPALQ-DVFVELLSPHSKRGKRGKADHTPEHSPST 537
QY 1494 NFRVATKSSQGRSVNTASKLSQESKRCRKRQSPSPVTLGRSSGRQGVHEL 1553
Db 538 NFRVSTRSRRLIPANTAESLSQHSRKRGRKQSTESSPVLNRRSSGRQGVHEL 597
QY 1554 SAFEOLVVELVRHDSWPFLKLSKIQVDPYDIKIPALNIIRKYNKCEYKLASEFI 1613
Db 598 SAFEOLVVELVRHDSWPFLKLSKIQVDPYDIKIPALNIIRKYNKCEYKLASEFI 657
QY 1614 DDIELMFCNCFEYNPRTSEAKAGTRLOAFHIOAQKLGHLVTPSNVDQVSTPPAAKSR 1673
Db 658 DDIELMFCNCFEYNPRTSEAKAGTRLOAFHIOAQKLGHLVTPSNVDQVSTPPAAKSR 717
QY 1674 I 1674
Db 718 I 718

```

RESULT 4

```

BAIA_XENLA
ID BAJA_XENLA STANDARD; PRT; 627 AA.
AC Q8UVR5;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE ATP-utilizing chromatin assembly and remodely factor 1 (XACF1)
DN (Fragment).
GN Name=XACF1;
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21668941; PubMed=11809820; DOI=10.1091/mbc.01-09-0441;
RA MacCallum D.E., Losada A., Kobayashi R., Hirano T.;
RT "ISWI remodeling complexes in Xenopus egg extracts: identification as

```

```

RT major chromosomal components that are regulated by INCENP-aurora B.";
RL Mol. Biol. Cell 13:25-39(2002).
CC -!- FUNCTION: May play a role in transcriptional regulation.
CC -!- SUBUNIT: Together with p18 and p20 proteins, it forms the xenopus
CC version of CHRAC.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- PTM: Phosphorylated in mitosis.
CC -!- SIMILARITY: Belongs to the WAL family.
CC -!- SIMILARITY: Contains 1 bromodomain.
CC -!- SIMILARITY: Contains 1 PHD-type zinc finger.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; AF412332; AAL60160.1; -.
CC HSSP; Q9UITG0; 1F62.
CC InterPro; IPR001487; Bromodomain.
CC InterPro; IPR011011; FYVE PHD_Znf.
CC InterPro; IPR001965; Znf_PHD.
CC Pfam; PF00439; Bromodomain; 1.
CC Pfam; PF00628; PHD; 1.
CC PRINTS; PR00503; BROMODOMAIN.
CC SMART; SM00297; BROMO; 1.
CC SMART; SM00249; PHD; 1.
CC PROSITE; PS00633; BROMODOMAIN 1; 1.
CC PROSITE; PS00104; BROMODOMAIN 2; 1.
CC PROSITE; PS01359; ZF_PHD 1; 1.
CC PROSITE; PS00016; ZF_PHD 2; 1.
CC KX Bromodomain; Coiled coil; Nuclear protein; Phosphorylation;
CC Transcription regulation; Zinc-finger.
CC FT NON TER 1
CC FT ZN FING 222 272 PHD-type.
CC FT DOMAIN 518 588 Bromodomain.
CC FT DOMAIN 281 327 Coiled coil (Potential).
CC SQ SEQUENCE 627 AA; 70208 MW; 42C17587E531D380 CRC64;

```

Query Match 19.6%; Score 1691; DB 1; Length 627;

Best Local Similarity 55.7%; Pred. No. 5.7e-62;

Matches 361; Conservative 90; Mismatches 157; Indels 40; Gaps 14;

QY 1044 KSRICAOQLARSEBEKHFESDKPQP--DSKPTYSGRSSNAVDPQSCAQLKLELRLRDL 1101
 Db 1 KSKICQRLDSPVNRNFNIADNTPPTLDFKP--FRGKLLNASESGFVSAEKQLELRLCDLI 58

QY 1102 LDIEDRIYQGTGLGAIKVTDRHWRSALESGRYELLSEENKENGIIKTWNEDVEEMEDQ 1161
 Db 59 LDIEDRIYQGTGLGAVKVPDRQLWRSALENNTLELNDSEKENGSSKPINHEWEEMED-- 116

QY 1162 TKVTVKDRLLGIKTETSTVSTNASTQSVSSVHYLAMALFOIEQGIERRFLKAPL-DA 1220
 Db 117 IKQSAKDRLLGLKNDTPSATSTSTNTQPNNVAVRYLARALHQIEQGVKRYLKAPLGA 176

QY 1221 SDSGRSYKTVLDRWRRESLLSASLSOVFLHLSTLDRSVIWSKSIILNARCKICRKGDAEN 1280
 Db 177 SESGRVQRTVLDRCGSELLSGSFSQIFLHLSTLDRSVIWSKSIILNARCKICRKGDAEN 236

QY 1281 MVLCDGCDRGHHTYCVRPKLKTVPEGDWFCPECRKPQRRLSPQRPSLESDEDDVEDSM 1340
 Db 237 MVLCDGCDRGHHTYCVRPKLKTVPEGDWFCPECRKPQRRLSPQRPSLESDEDDVEDSM 291

QY 1341 GGEDEVGDGDEEGQSEEEYEVQDEDDSQEE--EEVSLPKRGPQVRLPVKTRGKLSS 1398
 Db 292 --EEEELDQKEE--EEEEQEBELSENEQDEEMSESESPKRGRAKQVLPKLVKRGKAT 348

QY 1399 SFSRSGQQEPGRYPSPSQSTPTKTVSSKTGRSL-RKINSAPPTETKSLRIASRTPHS 1457
 Db 349 G-----XUGPKPKTGKOSTPKNTQPAPEGRGGKTKRSPASPLEPTSLRSASDSPAHG 400

735	EPMDVESDSEAP	PHFPELLMCSGDKRSCIVHDSRNGORQWAYIYKAEIDE	LIKALPN	794
1030	GHRESALKE	-----TLLOKSERIC-AQLARFSEKHFSDKPODQSKPTYSRGRSS	---	1079
795	GLREYELLQELS	VLRSLEIEHQHATCPVDLLSLENTWKKFMAAMESSTNRKYGEANFGL	854	
1080	NAYDPSQMAE	KOLELRRLRDLIDIEDRIYQGTILGAIKVTDRIHRSALYESGRYE	---	1134
855	PNGTDLNEV	-----MRLHLVDRITIOFENDIYVTGLGRLLKVKDMEKWRSDDLGGNYDAQCK	909	
1135	-----LJSEENKENG	ILIKTVNEDEVEMEIDBQTAKVIKD--RLGLIKTEPSTPSTVN	1184	
910	LQWPGGKLEDE	AGSDNESHETHEED--DGALLGKYARKPYRDPOMYLAASADTKPLPDS	968	
1185	ASTPQSVSV	-----VHYLAMALFQIEBQGIERRFLKAPLD-----ASDSGRS	1226	
969	DEEDQHTNVS	IPIAVHNMASSALLOVEQAIGKFLKEPYGMKKWDPKQEAALKACDSR	--	1026
1227	YKTVLDWRWS	LLSSASLSQVFLHLSTLDRSVIWSKSIILNARCKIKRKGDAENMVLCDG	1286	
1027	-----LHQWEVSL	MEBSTSAQVFLHLNLTHDCIQWRRSTNKSCLCKVCRRGSDPEKMLLCDE	1082	
1287	CDRGHTTYCVR	PKLKITYPEGDWCEPC-----APKORCRLSPRQPSLESDEBVE	1337	
1083	CNAGTHMFC	LKPKLRSYPPGNYWCNDCVKSLGLSNGQNEKDKQATKKRKFIVEED--	1140	
1338	DSMGGEDEV	DDEEQ-----SEEEYEVEDDDSQEEEV-SLPKGRPOVRLPVK	1391	
1141	-----DEATDEEE	EKKDDMTDEDAEHNEKHDEVEDDESVTSTPSSRVNGRLRR	1194	
1392	TRGLSSFSR	GOQOPGRYPSSQOSTPKTTVSSKTGRSLRKINSAPPTETKSRLIAS	1451	
1195	PRTRPTSR	RLTSKEIEB-----HAQEDVSDGVDDA--SLTAGEDTIEDSDEKVCQ	1246	
1452	R-----STRSHQ	PLQADVPEVLLS--PRKRGGRKSNANTPENS	1490	
1247	KCFYDGEIK	VCQRLFPFLHECVHLKRPPTDFVCKTCKEMPORRRRRHNMNGDHRDE	1306	
1491	NFNFRVIAT	KSSEQSRSVNIASKLSQESKXCRCKROSPEPSPVTLGRSSGQGV	1550	
1307	EEP-----KAKPR	NSLRLSDTKTARPSNGNNNNNNNS--SVNNNHRSGRGRTNE	1356	
1551	H-----ELSAF	EQLVVELVRHSDSWPFLKLVSKIQVPDYDIICKPIALNIIREKVNKECYK	1607	
1357	HMPLNSA	ALYDLLEQIMKHAAPFLRPVLTSVPDYHQIITKTPMDLAKIKSLNMGAYQ	1416	
1608	LASEFTDDI	EILMPSNCFEYNPRNTSEAKAGTRLQAFPHIOAKQLGLHVTSPNVD	1661	
1417	LNPELLASDI	OLAVFRNCILYNVEGNEIYDAGCOLERFVIDRCWOLPFRPSDMN	1470	

RESULT 8
O9Y0W1

Q910W1
ID Q9Y0W1
PRELIMINARY;
PRT: 1476 AA.

AC Q9Y0W1;

DT 01-NOV-1999 (TREMBLrel. 12, Created)

DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)

DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)

DE ATP-dependent chromatin assembly factor large subunit.

GN Name=Acf1;

OS *Drosophila melanogaster* (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.
OC

OX NCBI_TaxID=7227;
BN [1]

RN
 [I]
 CEQUENCE FROM N A

RP
SEQUENCE FROM N.A.
PY MEDLINE-99315527. PubMed-10395522.

RX MEDLINE=99315627; PubMed=10385622;
 RA Ito T Levenstein M E Evdorov D V Kutach A K Kobay

RA ILO I., LEVENISLETH M.E., FYODOROV D.V., KUTACH A.K., KOBAY
RA Kadonaga J.T.:

KA Naohaga J.T.;
RT "ACE consists of two subunits Acf1 and TSW that function

ACF consists of two subunits, ACT1 and ISW1, that function cooperatively in the ATP-dependent catalysis of chromatin

RI. Genes Dev. 13:1529-1539 (1999)

TY 81100-1333(1999).

[illegible]

```
Db 578 NRKLLTQMHLEVNGVEPEKREALVEKLK-----SIAE----- 611
Qy 840 QKELDQDMFTEDDDPGSHKGRGRKQNGFKBFTROEQINCVTRELLTADBEALKE 899
Db 612 -----LHAQSD-----QQ 619
Qy 900 HORKEKELLEKIOAIACTNIFPLGRDRMYRYWIFPSIPGLFTIEDYSGLTEDMLLPRP 959
Db 620 HRKHELQWL-KLHQLFNFLVY-LGMDRCYKRYVLESMPGIFVEHSPDSL--DTCLEQP 675
Qy 960 SSFQNNVQSDPQVSTKTGEPLM-----SESTSNIDQG 992
Db 676 ITNKSQIEIRQOSALPKNRKDLRVLLKLYGDDEKTKKAKHKSLENKENQERLNGSAE 735
Qy 993 PRDHSVQLPK-PVH-----KPNRWCFYSSCEQLDOLIEALNSRG 1030
Db 736 PMDVESNPEAPTHFELLMCSGDKRSCIVHDSRQQRWAYIYKABEIDELIKALNPNG 795
Qy 1031 HREGALKE-----TLQEKSRIC-AQALRPFSEKFPSPDKPQDPKPTYSRGRSS---- 1079
Db 796 LREVELLOELSVLRSLIEHQAKTCEVDLLSLENETMRKKFMAAASETNRYKGEANGLP 855
Qy 1080 NAYDPSQMAEKQLELRDLFLDIEDRIYQGTILGAIKVTDRIHWRSALSGRYE----- 1134
Db 856 NGTDLNEV-----MRLHLVDRIIQFENDIYTGDLGRLLKVKDMKWRSDLLGNGYDAQCKL 910
Qy 1135 -----LLSENKENGIIKTVNEDVEEMEDQTKVIVKD--RLLGKHTTPTSTVSTNA 1185
Db 911 QWPGGKLEDAAGSDNESHETHEED-DGALLGKVKARPYRDPGMYLAASADTKLPDPSDD 969
Qy 1186 STPOSVSV-----VHYLAMALFOLEOGIERRELKAPLD-----ASDSGRSY 1227
Db 970 BEDQHTNAVLPIAVHNNASALLQVEQAIGKRFLEPEYGMKKWDPKQALKKACDSR--- 1026
Qy 1228 KTVLDNRRESLLSSASLSQVFLHLSLTDLSRVSIVSKSILNARCKCRKKGDAENNVLDGCG 1287
Db 1027 ---LHQMVEVLSWESTSPAQVFLHLNLHDCIQWRSTNKSCLKVCRGSDPEKMLLCDEC 1083
Qy 1288 DRGHHTYCVRPKLKTVPEGDFWCFEC-----RPQRCRRISFRORPSPLESDEVED 1338
Db 1084 NAGTHMFLCKPLRSVPPGNNYCNDCVKSLGSLHGQNEKDKKQATKKRKFIVEED--- 1140
Qy 1339 SMGGEDDVEDGDEEGO-----SSEEEVEVEQDEDDSOEEEV-SLPKRGPRQVRLPKVT 1392
Db 1141 -----DEATDEEBEKKDDMTDAEHENKHDVEDDEDSVTSTPSSRVNGRIILRRP 1195
Qy 1393 RGLSSFSFGQOQEPGRYPSRQOSTPKTTVSSKTSGRSLRKINSAPPTETKSLRIASR 1452
Db 1196 RTRPTSRRLTSKETEE-----HAQEDVSDGVSDDA--SLTAGEDTIEDSEDEKVCQK 1247
Qy 1453 -----STRHSHGPLQADVVELLS--PRRKRGRKSANNTPENSPN 1491
Db 1248 CFYDGEIKVCQCLFFHLECVHLKRPRTDFVCKTCKPMPQRRRRHSNMGDHRDEE 1307
Qy 1492 FPNFRVIATKSSQSRVNIASKLSQSESKRCRKQSPPEPSPVTLGRSSGRQGVH 1551
Db 1308 EP-----KAKRPNLSURLSIDKTARPSNGNNNNNNNS-----SVNNNNHRSRGRRTNEH 1357
Qy 1552 ---ELSAPEQLVBLVRHDDSWPFLKLVSKIQVDPYDIIKKPALTALNIIREKVNKCYKL 1608
Db 1358 MPLNSAALYDLLEQIMKHKAAPFLRPVLTVSEVPDYHQIITKTPMDLAKIKSLNNGAYQL 1417
Qy 1609 ASEFIDDIELMFSNCFEYNPRNTSEAKAGTQLQAFPHIOAQKGLGHTVPSNVD 1661
Db 1418 NEELLSDIQLVFRNCDLYNVEGNEIYDAGCQLERFVIDRCRDMQLFRPSPDMN 1470

RESULT 9
Q9NG24 ID Q9NG24 PRELIMINARY; PRT; 1476 AA.
AC Q9NG24;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
```

```
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Chromatin accessibility complex (CHAC).
GN Name=Acfl;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21340270; PubMed=11447119; DOI=10.1093/emboj/20.14.3781;
RA Eberhart A., Ferrari S., Langst G., Straub T., Imhof A.,
RA Varga-Weisz P., Wilim M., Becker P.B.;
RT "Acfl, the largest subunit of CHAC, regulates ISWI-induced nucleosome
remodelling.";
RL EMBO J. 20:3781-3788 (2001).
DR EMBL; AJ238397; CAB88669.1; -.
DR HSP; Q9UIG0; 1F62.
DR FlyBase; FBgn0037620; Acfl.
DR GO; GO:0016590; C:ACF complex; IDA.
DR GO; GO:0008623; C:chromatin accessibility complex; IPI.
DR GO; GO:0003677; F:DNA binding; TAS.
DR GO; GO:0042766; P:nucleosome mobilization; IDA.
DR GO; GO:0016584; P:nucleosome spacing; IDA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IDA.
DR InterPro; IPR001487; Bromodomain.
DR InterPro; IPR004022; DDT.
DR InterPro; IPR001965; Znf PHD.
DR Pfam; PF00439; Bromodomain1.
DR Pfam; PF00628; PHD; 2.
DR SMART; SM00297; BROMO; 1.
DR SMART; SM00571; DDT; 1.
DR SMART; SM00249; PHD; 2.
DR PROSITE; PS50014; BROMODOMAIN_2; 1.
DR PROSITE; PS0827; DDT; 1.
DR PROSITE; PS01359; ZF PHD_1; 2.
DR PROSITE; PS50016; ZF_PHD_2; 1.
SQ SEQUENCE 1476 AA; 170469 MW; F9B6FD1F1EF8D235 CRC64;

Query Match 14.8%; Score 1275.5; DB 2; Length 1476;
Best Local Similarity 24.3%; Pred. No. 1.9e-44;
Matches 407; Conservative 283; Mismatches 617; Indels 367; Gaps 46;

Qy 151 MPLLHRKPP-VROKPPAD--LRPDEVPYCKVTNEIFRHYDDPFERILCNLSWCAVT 207
Db 1 MPICKREGFDLNQEGKNETFHDNDQVFCCYITKRIFRDYEHYPRHVMINSTVQCEAT 60
Qy 208 GRPLGTYQEALESEKKARONLQSPPEPLIIPVLVLTSLTHRSRLHEICDDIFAVVKDQRYF 267
Db 61 GKENLTVEEAIVKSRARKQWQEQKQSLRSPVLLVVEHAQOASAVTNLMIVAKELRKYP 120
Qy 268 VEETVEVIRNNGARLQCTILEVLPPSHQNGFANGHVNSVDGETIIISDSDSETQSCSFQ 327
Db 121 IGEVSVQAKKNATY--TVLGVKLDKNWPEPLNG-----IYEDTDN----- 159
Qy 328 NGKKDAIDPLLFYKVVQPTKKELHESAIYKATQISRKKHLFSRDKLKLKQHCPEQBG 387
Db 160 -----LVYRLRPNKGDPSAELDLFPFQRLRRSRMEFNLENLSMFIKSNVSRVDG 207
Qy 388 VIKIKASSLSYTKIAEQ--DPSYFFPDDPPTFFISPANRRGRPPKRIHISQEDNVANKQ 445
Db 208 LLRKPAPYKQY-VTDGPNVFSTIFIGKMPR--YSPAKIKKPGQKKQ-----STLNKY 257
Qy 446 TLASYRSKATKERDKLLKQBEKMSLAFKAKLKREKADALEAKKKEKEDKKEELKKI 505
Db 258 IVA---GEATAAKSKAKSDAKSLAEELERVKREK---EAKLIIELE----- 298
Qy 506 VEEERLKKKEERLKVREKEREKLEERKRVVEYLYKQWSKPEDMECDLDELPEPTP 565
Db 299 -----KQAEKKAQLIERVENECNLLQK-----TDDLERTDQKVLPRYRQ 339
```

Qy	566	VKTRLPPIFGDALVMVLEFLNAFGLFDLODFPDGVTVLEVLEBALVGNDSBGLCLBLFF	625
Db	340	IVTLLPEHLGGDAFMWREBMHTYTGLLSGIEVFVRQLNSFYEMTRALTARETAGSLDILL	399
Qy	626	FFLTAIFAIAEEEEE-----VAKEQLTDAUTKGCSLKSLDLS	664
Db	400	VLLGTVDFOKEEBEECNAVTYLDRAAOTQBPYPWSMAQAASHLYAKHHFSFKVNELPDA	459
Qy	665	CTLSEILRLHILASGADVTSANAKRYQKRGGFDATDDACMELRSLNPSVLKKLSSTSVY	724
Db	460	LTLNEVLRLHLGSGAFVNEKAERWMYRNGYSKEDPGLELRLESHILIRILKNHSVY	519
Qy	725	DLTGECKMIHALCGKLLTLVSTRDFIEDYDILROAKOEFRELKAEQHRKEREREAAAAR	784
Db	520	QLKFCDMLLRCLMSQITMYSGTINLTIERMEQTAKARODRALVALVGXENR---LAAVE	576
Qy	785	IRKEKEEKIKEOE-----QMKKEQEKJLEDORNSTADISICEBEREDFTSIESKDT	838
Db	577	INRKULTQMHHLEVNGVEPEKREALVKLK------SIAB-----	611
Qy	839	EQBELDQDMFTDEDDPGSHRRGRGRKGQGFEXFEFTRQEQCINCVRELLTDAEBEALKQ	898
Db	612	-----LHAQSD-----Q	618
Qy	999	EHQRKEKELLEKIOAIACTNIFFPLGRDRMYRRYWIFFSPICGLFIEDYSGLTEDMLLPR	958
Db	619	QHRKHHELQML-KLHSOLFNFVLV-LGMDCRCYKYVYLSEMPGIFVEHSPDSL--DTCLEQ	674
Qy	959	PSSFQNNVQSOPQOVSYKTGEPLM-----SSTSNIDQ	991
Db	675	PITKNSQIETRQQSALPKNRKDRLVYLLKYGDDKTKKAKHSLKENQEHRLNGSA	734
Qy	992	GPRDHSVOLPK-PVH-----KPNRWCFYSSCEQLDOLIEALNSR	1029
Db	735	EPMDVESDSPAPTHPELLMSCGDKRSQIVHDSRNGORQORWAYIYKAEIDELIKALNP	794
Qy	1030	GHRSEALKE-----TLLOEKSRTC-AQLARFSEEKPHFDKPODPDSKPTYSRGRSS---	1079
Db	795	GLREYELLQELSVLRSILTEQHAKTCPVDLLSLENETWRKKFMAAMESETNRKYGEANFGL	854
Qy	1080	NAYDPQMCAEKOLELRDPFLDIEDRIYQGTLGAUKTDRHWRESALESGRYE----	1134
Db	855	PNGYDLNEV-----MRHLVDRIIOFENDIYTDGLRLKVYKDKMEKRSDDLGGNYDAQCK	909
Qy	1135	-----LLSEENKENGIIKTVNDEVBEIMBDQTVIVKD--RLGIKITEPTSTVSTN	1184
Db	910	LOWPGCKLEADGSNDSESHETHEED-DGALLGKYARKPYRDPGWYLAASADTKPLPDS	968
Qy	1185	ASTPOSVVV-----VHYLAWALFOIBEGIERREFLKAPLD-----ASDSGRS	1226
Db	969	DEEDQHTNAVSIPITAVHNMASALLQVEQAIGKRFLEKPYGMKWDPKQEAALKKLACDSR--	1026
Qy	1227	YKTVLDRWRSSLSSALSQVFLHLSTLDRSVIWSKSILNARCKICRKGDAENMWLCDG	1286
Db	1027	----LHOWEVSLMESTSFAQVFLHLNLHDICIOWRRRTNKSLCKVCRCRGSDPEKMLLCDE	1082
Qy	1287	CDRHHTYCVRPKLKTYPEGDWFCPEC-----RPKCRRLSPROPSPLESDEBVE	1337
Db	1083	CNAGTHMFCLKPLRSVPPPGNWCUCVKSILGNGNEKDKKOATKKKRFIVEBED--	1140
Qy	1338	DSMGEDDEVGDDEEQ-----SEBEEYEVEDQDDDSQBEEEE-SLPFRGRFPQVRLPVK	1391
Db	1141	-----DEATDEEBEEKDDMTDEDAEHENEKHDEDVEDDESVTSTPFSSRVNGRIIRR	1194
Qy	1392	TRGLSSFSRGOQQPGRYPBSRQOSTPKTVTSSKTGRSLRKINSAPTETYSURIAS	1451
Db	1195	PRTPTRSRLTSKEIE-----HAQEDVDSGDVSDDA--SLTAGEDTIEDESDEEKVCQ	1246
Qy	1452	R-----STRSHCHPLQADVVELLS--PRKRGKRSANNTPNSP	1490
Db	1247	KCFYDGGEIKVCQCRRLFHLECVHLKRPRTDFVCKTKCMPORPRRRHHNMGMGDHORDE	1306
Qy	1491	NFPNFRVIATKSSSQRSRVNIASKLSIQESSESKRCRKROSQPSPSYTLTGRRSSGROGV	1550

```

Db      1307 EEP-----KAKRPRNSLRLC|DKTARPSNGNNNNNNNS|-----SVNNNHRRSGRRTNE 1356
Qy      1551 H--E--LSSAFQLVVELVRHDDSPFLK|VSKI|OVDPYD|I|KKPIALNI|IREKNVCKEYK 1607
Db      1357 HMP|NSA|LYDL|EQ|MKHKA|MP|LPV|LT|SE|VPD|YH|Q|I|KTPMD|AK|IKSLN|MGAYQ 1416
Qy      1608 LASEFIDDI|ELMF|SNCFEY|NP|RT|SEAK|G|TR|LQA|FFH|Q|A|Q|L|GLH|VT|PS|NVD 1661
Db      1417 LNE|LLSD|I|Q|V|RN|CD|LY|N|VEG|NE|YD|AC|QLER|FV|ID|CR|DM|QL|P|FR|SD|WN 1470

RESULT 10
Q7QBK2
ID Q7QBK2 PRELIMINARY; PRT; 1202 AA.
AC Q7QBK2:
DT 01-MAR-2004 (TREM|B|rel. 26, Created)
DT 01-MAR-2004 (TREM|B|rel. 26, Last sequence update)
DT 01-MAR-2004 (TREM|B|rel. 26, Last annotation update)
DE AGCP2918 (Fragment).
GN Name=agCG52496; ORFNames=ENSGG00000019007;
OS Anopheles gambiae str. PBST.
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Anopheles.
OX NCBI_TaxId=180454;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PBST;
RA Anopheles Genome Sequencing Consortium;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
CC -! CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AAB01008979; EAA08347.1; -.
DR HSP; Q9UIG0; Irf62.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR004022; DDT.
DR InterPro; IPR011011; FYVE PHD_ZnF.
DR InterPro; IPR001965; ZnF_PHD.
DR Pfam; PF00628; PHD; 1.
DR PROSITE; PS50827; DDT; 1.
DR PROSITE; PS50016; ZF_PHD_2; 1.
FT NON TER 1202 1202
SQ SEQUENCE 1202 AA; 134820 MW; A869AB9ACD57C529 CRC64;

Query Match 14.2%; Score 1228; DB 2; Length 1202;
Best Local Similarity 25.2%; Pred. No. 1.4e-42;
Matches 349; Conservative 215; Mismatches 415; Indels 408; Gaps 33;

Qy      151 M|PL|LR|K|P|F|V|R|Q|P|P|A|D|L|R|P|D|E|V|F|V|C|K|T|N|E|I|F|R|H|Y|D|D|F|F|T|I|C|N|S|L|V|W|S|C|A|V|T|G|R|P 210
Db      1 M|PL|K|R|K|P|L|Q|V|P|D|Q|E|R|K|D|G|E|V|F|V|C|T|T|G|E|L|F|S|Y|D|D|F|F|N|K|L|S|L|T|V|W|S|C|A|V|T|G|R|S 60
Qy      211 G|L|T|Y|Q|A|L|E|S|E|K|K|R|Q|N|L|Q|S|F|P|E|P|L|I|P|V|L|Y|L|T|S|L|T|R|S|R|L|H|E|I|C|D|D|I|F|A|Y|V|K|D|R|F|V|E|E 270
Db      61 N|L|Y|A|D|A|L|E|S|E|K|A|K|R|L|K|S|F|P|T|A|L|G|P|I|L|I|A|S|R|K|R|T|A|I|H|E|L|A|S|D|V|H|G|A|K|D|V|F|K|G|E 120
Qy      271 T|V|-----E|V|R|N|R|G|A|R|Q|C|T|L|E|V|L|P|S|H|Q|R|F|G|A|R|H|V|N|S|V|D|G|T|I|I|S|D|S|D|D|S|E|T|Q|S 323
Db      121 T|V|T|K|A|A|D|S|E|T|V|R|K|-----G|K|I|V|R|V|L|G|E|L|P|S|D|H|Q|----- 151
Qy      324 C|S|F|Q|R|G|K|K|D|A|I|D|P|L|P|K|Y|Q|V|P|T|K|E|L|H|E|S|A|I|V|K|A|T|Q|I|S|R|R|K|H|L|F|S|R|D|K|L|K|L|P|L|K|O|H|C|E 363
Db      152 -----P|S|R|L|Y|Q|I|V|G|S|D|D|E|T|P|A|S|Y|T|R|G|D|A|V|T|R|E|R|N|C|L|S|R|E|K|R|C|K|L|F|L|K|O|H|V|E 198
Qy      384 --P|Q|E|G|I|K|I|K|A|S|L|S|T|Y|K|I|A|E|---Q|D|F|S|Y|F|F|D|D|P|P|T|F|I|F|S|P|A|R|R|R|R|G|R|P|P|K|R|I|H|S|Q|E|D 439
Db      199 L|G|P|D|Q|---V|L|C|V|K|K|A|L|S|Q|F|V|T|S|K|G|C|T|D|K|V|F|Y|G|T|P|D|G|V|-----S|K|L|---L|A|K|E|T 245
Qy      440 N|V|A|N|K|O|T|L|A|S|R|K|A|T|K|R|D|K|L|L|Q|E|M|K|S|L|A|F|E|K|A|K|I|K|R|E|K|A|D|A|L|E|A|K|K|E|D|K|E|K|R 499
Db      246 T|M|S|R|K|L|L|S|Y|V|A|N|A|K|Q|Y|D|S|I|----- 267

```


Qy 500 BELKIVVEERLKKKEKERLKVREKER EKLREKRYVEYLKQWSKPRDEMECDLKE 559
Db 268 -----HEQQLMQRV 278
Qy 560 LPEPTPVKTRUPPEIFGDALMVLBNFAGBELFDLQDEFPDGVTVLEVLEALVNDSEGP 619
Db 279 LPPARPVQTGLTEHFFSFMFLBFLHQFDLSIEBKFPDGVTVLEQLEASRAVNGP 338
Qy 620 LCELLFPLFTAIFOAIAEBEBEVAKEQLTDADTK-----CSSLKSLDLDSCTLSEI 670
Db 339 LSDIFQVLLCTLFAAL-KESHSSBSNLPCATLVRVWKCRKHFGAKUTLDPMDSTTVTEL 397
Qy 671 LRLHLASGADVTSANAKRYQKRGPGDATDDACWELRLSNPSLVKKLSSTSVYDLTPGE 730
Db 398 LRLHLVAYGQP-----NHFGDACFTLIRHPQLMRLTHTVTVQLPTAI 441
Qy 731 KMKILHALCGKGLTLTVSTRDFIEDYVDILROAKQEFRELKAEQHKEREAAAIRKRKE 790
Db 442 VLQVLCALVHQLQL-----DEVLERVEKVG 468
Qy 791 EKLKEQEKMEKEKELKEDQRNSTADISGEEREDFDTISKEKTEQKELDQDMPTE 850
Db 469 ARAKFNSNRTAQRQ-----LAKRTASLKQAQADQGVTR 500
Qy 851 D-----EDDPGSHKGRGRKRGQNGKFEFTROEQINCVTRRELLTADDEEALKQEHQRK 903
Db 501 EMASLMVGDQPADSAKVEKR-----ARLQQ--KLQELAQIADASQLAELO 548
Qy 904 EKELLEKIQSAIACTNIFPLGRDRMYRYVFPISPLGFIRED----- 946
Db 549 DQ--YEGNLNLSYGLYEVY-LGSDAFRNYMKFESLPLGLFVEHDGTFAGCCSERVTPHLP 605
Qy 947 -----YSLTEDMLLPSPSQNVV----- 966
Db 606 LVACHPKRLKRYITYAILKCAAGNSDGLDGLVLAGRQEQKSDLCQLLVLRGIALQKEEA 665
Qy 967 QSODPOVSTKTEPLMSBESTNIDQGRDHSVOLPKP----- 1003
Db 666 RKKNPFLAKAGE-----NASVAEDTADNMWIDEFDSSVPPPAAPTRELLMCTGNP 719
Qy 1004 -----VHKPNR-----WCFFSYSCQLDLQLEALNSRGHRESALKETLLQEKSRICAQLA 1052
Db 720 EDCVTHIASROTAPGTTWSYATADELDALIASLNERGLREKSLRKTLEQYRDGIVKRLK 779
Qy 1053 RFSEKEHFDKPODPSKPTYSRGRSSNAYD--PSQWCAEQQLRLRDLFLDIEDRIYQ 1110
Db 780 KCPKIS--LSHRQPSAAPARAQHRQCSAEFPVPEAVTNETLEMFRDQLLMEGRHA 837
Qy 1111 GTLGAIKVTDRIHWSALESGRYELLSEENKENGIIK-----TVNED--V 1153
Db 838 GCLGELKVNSTEKRIALNRSYDAQITGQLQWVRRRQRLEEMVYDKLDAEDSSSESGV 897
Qy 1154 EMEIDEQTKVI-----VKDRLLGKIKTETPTSTVSTNASTPOSVSVSHYLAMALFQI 1205
Db 898 BETEVADEIDLAKHRADPGYADLYSAERSENGTDGMALETSDEIQKTVHSLACALLQV 957
Qy 1206 EQGTERPLKAPLD-----ASDSGRSYKTVLDLRWRESLLSSASLSQVFIHLSTL 1254
Db 958 AQSIETPKRLRFPFGKGRCKDRNTVAMQFRGQKRLLD--WEGSLMRATSYSQLFLHYHIL 1016
Qy 1255 DRSVIWSKSLNARCKIKRKGDAENMVLCDGCRGHHTYCVRPKLTVPEDGMFCPECR 1314
Db 1017 YDALYWSKSIERAVCMVCRKRGDANLTLDCNCRACHMYCLPKLKKVPEGDMFCMKCR 1076
Qy 1315 PKQRCRLRSFRQPSLE-----SDEVEDSMGEGEDVDGDEEGQSEEEVEVEODE 1367
Db 1077 PSDDAGIVSRKRTQTIEDLDLSDNDSSAATNESEDD--BEERKEEGEEDELEMESED 1135
Qy 1368 DDSQEE 1374
Db 1136 STTSESE 1142

RESULT 11

Q66IE0 PRELIMINARY; PRT; 371 AA.
AC Q66IE0
DT 25-OCT-2004 (TREMBlrel. 28, Created)
DT 25-OCT-2004 (TREMBlrel. 28, Last sequence update)
DT 25-OCT-2004 (TREMBlrel. 28, Last annotation update)
DE Hypothetical protein (Fragment).
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Olfactory epithelium;
RX PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Whiting J., Helton E., Kettman M., Madan A.C., Rodriguez S., Sanchez A.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield V.S.,
RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RN and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Olfactory epithelium;
RA Director MGC Project;
RL Submitted (SEP-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC081390; AAH81390.1; -
DR InterPro: IPR00637; A+T.hook.
DR Pfam: PF02178; A+T.hook; I.
KW Hypothetical protein.
FT NON TER 371
SQ SEQUENCE 371 AA; 43339 MW; 4C6D7689428901B6 CRC64;
Query Match 13.8%; Score 1194; DB 2; Length 371;
Best Local Similarity 59.8%; Pred. No. 9.3e-42;
Matches 230; Conservative 69; Mismatches 68; Indels 24; Gaps 7;
Qy 151 MPLLHRFPFVRQKPPADLRPDEEVFYCKVTNEIPRHYDDFFERTILCNLSWCAVTCRP 210
Db 1 MPLLHKFPFVRQKPPADLRPDEEVFLCKYTHIEFTYDEFFERTILCNLSWCAVTKGP 60
Qy 211 GLTYQEALESEKKARQNLQSFPEPLIIPVLVLTSLTHRSRLHEICDDIFAYVKDRYFVEE 270
Db 61 GLTYQEALESEKKARLSQLQNFNALAVPLLHLTALTALHRSRLHEICDQVAYVKDRFPFGE 120
Qy 271 TVEVIRNNGARLQCTILEVLPPSHQNGFANGHV-NSVDGETIILISDDSDSETQSC--SFQ 327
Db 121 MVDITSSTWQL-CKILEVIPP-YSGNAGVNGHTKRIEGDSIVISDSDDETTSPTKTPS 178
Qy 328 NGKKKDAIDPLLKFKYQVQPKKELHESAIVKATQISRKKHLFSRDKLKLFLKQCEPOEG 387
Db 179 NGKKKKAINESSFKYRQVPVSPEACEFPIVKVQMSRKKSIFSRDLKLLKQCEPMHG 238
Qy 388 VIKTKASSLSTYKIAEQDFYFPDPTTFFSPANRRRRGPPKRIHISQEDNVANKQTL 447
Db 239 TIRVKPSSALQYKLAHSFSHFPPDEPPVFTFSPGKGRGPPPNNASSTEINYI----- 292

Db 504 QFSIIIAKCNDPQNSFADLLRVLLSIRTDIAEDGDEADINNREEVYLINAQNCDP 563
Qy 643 -----AKEOLTDADTK-----GCSLKSILDLDSCTILSEILRLHLASGADVTANAKYR 690
Db 564 AHVTHGSDIRLSDLHFIRKIRKIGKSVHRLPVDWMTLTVLRLFTFSGYTGWATHRRH 623
Qy 691 YQKGGGFATDACCWELRLNPSLVKLSSTSVYDLTPGEXKWKILHALCGKLLTLVSTRD 750
Db 624 LYARGNFRGYEDPAYEFTRHPGIMEKRLTUTVFDLEAPERLEIVKTLIYQLLYTSKFRG 683
Qy 751 FIEYVDILROAKQEFRELKAEQHRKEREAAAIRKR-----KEEKLBQEQ 798
Db 684 HLEQRONELVELKREQKKAWDVQGEAANAARLLELAPAGASGAPDVKEQKEPPVVR 743
Qy 799 KWKKEQKLEKDEQBNSTADISIGEEREDPDTSI-ESKOTEQKELQDMFTEDEDDPGS 857
Db 744 RLKAHAKAINEGR-----YKEDLDIAVLESVPVYASLSLD----- 779
Qy 858 HKRGRGRKGQNGFKEPTRQEQINCVTRRELLTABEEALKQEHORKEKELLEKIOSAIAC 917
Db 780 -----EIVTA--RELQKSEFKILMDSLISKMFQIYSK 809
Qy 918 TNPFLGRDRMYRWIPSPIGLFIE-----DYSGLTEDMLLPSPFQNNVQ 967
Db 810 ISDIRLGSDRAYRIVMENLSAILVETATFQELGIYCDPSIIDPL-----NLL 861
Qy 968 SQDQOVSTKTCEPLMSBSTSNI DOGPRDHSVOLPKPVHKNRWCIFYSCCEQLDOLIEALN 1027
Db 862 NEHQEVFTCTG-----NMDT-CRVHGD-----SNNWTRWSYIRDREQFQELLSLN 907
Qy 1028 SRGHRESALKETL-----LQKSRI CAQLARSEEFKPE-SDKPOQDSPKTYTSRG 1076
Db 908 PRGNREVELLEELNEYRPSLEILEETRLHEEDEDKWSQMTWDPNDPGD--TYN-- 963
Qy 1077 RSSNAYDPSQWCAEQLELRDLFLDIEDRIYQGTGAIK---VTRDHIWRSALSGRY 1133
Db 964 -----IDWDAEMRDLDDLEEKIDQOGMSIEKIFECNRIEWRDL----- 1004
Qy 1134 ELLSEENKENG-IKTVNEDVEEMEIDEQTKVIVKDRLLGLTKTPTSTVSTNASTPOSVS 1192
Db 1005 -----KESGNVCMLLNEDIS-----IFG-----EISYNLESESHFS 1035
Qy 1193 SVVHYLAWALFOIEQGTERRFLKAPLDA--SDSGRSYK--TVLDRWRESLSSASLSOVF 1248
Db 1036 DSMK-LAIAFWIYIKSHLKFIKAPYSPNKDEHGNLKPSELFRWQALLECESHALS 1094
Qy 1249 LHLSTLDRSVTWSKSIILNARCKIRKKGDAENVMVLCDCDGRGHYTCVVRPKLTVTPGDW 1308
Db 1095 LP1STFEGSIKWDSRLQKCRSKRRAAAHDLVLCSECDNCYHLKCAKLDVNSDAPADW 1154
Qy 1309 FCPCEPRKQRCRLSFRQRPSEDEDEVDGSGGDEDEVDGEGQSEEEYEVEQDED 1368
Db 1155 MCTSCRAQOR-----KVENEAKKMARDDDD 1178
Qy 1369 ---DSQEEVSLPKRGRPOVRLPVKTRGKLSSSFSS--RGQQPQPGRPYPSRSQOSTPKTT 1424
Db 1179 QLAESSQEDVAIGNI-----SISDNFSETAPTAEPASCSPTTTSKEPPT 1225
Qy 1425 VSSGTGSLRKINSAPPTETKSLRIASRSHSHGPLQADVVELLSPRRKRRGRKGSANN 1494
Db 1256 IRTASGRAVKRVHYSDIIEGLSL-----KMKRKSNGS 1256
Qy 1485 TPENSPNPNFRVATKSEOSRSVNI-----ASKLSQESKSKRCRKRQSPSPSPV 1537
Db 1257 NP-----VTPSERPVRNVSRIFDVENENVLTDDEDDSGENTRKRKPTSKKS 1304
Qy 1538 TLGRSSSGQGV-----HLSAFEQIAVLVLRHDDSWPFLKLVSKIQVDDYDIITKPP 1591
Db 1305 VTSTPTTNDISRVIPNIKEKKWTLEITLKKAMQECSSWPFLQPVDSKEVDDYDVIKRP 1364
Qy 1592 IALMIIREKVNKCYKLASEFIDDIELMFNSCFYNPNTSEAKAGTRLQAF 1643
Db 1365 MNLRTMWNKIQRINYKNPIERVNDQFLILSNCTYNEPENEIYKLSRELHDF 1416

RESULT 13

BA1B MOUSE
ID BA1B_MOUSE STANDARD; PRT; 1479 AA.
AC Q92277; Q9CU68;
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Bromodomain adjacent to zinc finger domain protein 1B (Williams-Beuren syndrome chromosome region 9 protein homolog) (WBSR9).
GN Name=Baz1b; Synonyms=Wbscr9;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9907764; PubMed=9859827;
RA Peoples R.J., Ciasco M.J., Kaplan P., Francke U.;
RT Identification of the WBSR9 gene, encoding a novel transcriptional
regulator, in the Williams-Beuren syndrome deletion at 7q11.23;
RL Cytogenet. Cell Genet. 82:238-246(1998).
RN [2]
RP SEQUENCE OF 1-482 FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Thymus;
RX MEDLINE=22354683; PubMed=12466851; DOI=10.1038/nature01266;
RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
RA Nikaido I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,
RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,
RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
RA Schrell L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,
RA Blake J.A., Bradt D., Bruscia V., Chothia C., Corbani L.E., Cousins S.,
RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazee J.S.,
RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,
RA Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,
RA Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,
RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
RA Meglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,
RA Nagashima T., Numata K., Okido T., Pavan W.J., Perlea G., Pesole G.,
RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramchandran S.,
RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,
RA Sandelin A., Schneider C., Semple C.A., Setou M., Shimada K.,
RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,
RA Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,
RA Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang L., Yang L.,
RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,
RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,
RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,
RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,
RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,
RA Yaeunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
RA Birney E., Hayashizaki Y.;
RT "Analysis of the mouse transcriptome based on functional annotation of
60,770 full-length cDNAs";
RL Nature 420:563-573(2002).
CC -!- FUNCTION: Forms a chromatin remodeling complex that mobilizes
nucleosomes and reconfigures irregular chromatin to a regular
nucleosomal array structure (By similarity).
CC -!- SUBUNIT: Interacts with ISWI (limitation SWI protein) to form the
WTF-ISWI chromatin remodeling complex (WICH) (By similarity).
CC -!- SUBCELLULAR LOCATION: Nuclear (Potential). Accumulates in
pericentromeric heterochromatin during replication (By
similarity).
CC -!- TISSUE SPECIFICITY: Expressed in all tissues examined including
heart, brain, spleen, lung, skeletal muscle, kidney and testis.
CC -!- DEVELOPMENTAL STAGE: Expressed as early as day 7 and in equal
amounts during gestation.
CC -!- SIMILARITY: Belongs to the WAF family.
CC -!- SIMILARITY: Contains 1 bromodomain.
CC -!- SIMILARITY: Contains 1 DPT domain.
CC -!- SIMILARITY: Contains 1 PHD-type zinc finger.
CC -!- SIMILARITY: Contains 1 WAC domain.

FT DOMAIN 850 893 Coiled coil (Potential).
 FT DOMAIN 1245 1283 Coiled coil (Potential).
 FT DOMAIN 1261 1273 Poly-Glu.
 FT VARSPLIC 660 663 Missing (in isoform 2).
 FT 14 14 /FTID=VSP_000552.
 FT CONFLICT 14 14 K -> N (in Ref. 3).
 FT CONFLICT 22 22 L -> P (in Ref. 3).
 FT CONFLICT 136 136 K -> E (in Ref. 1).
 FT CONFLICT 191 191 Y -> D (in Ref. 4).
 FT CONFLICT 298 298 Y -> V (in Ref. 4).
 FT CONFLICT 823 823 E -> R (in Ref. 3).
 FT CONFLICT 1191 1191 R -> P (in Ref. 3).
 FT CONFLICT 1354 1354 K -> M (in Ref. 2).
 FT CONFLICT 1438 1438 A -> V (in Ref. 3).
 SQ SEQUENCE 1483 AA; 170902 MW; 0CC146FEBB954261 CRC64;

Query Match
 Best Local Similarity 10.9%; Score 941.5; DB 1; Length 1483;
 Matches 399; Conservative 263; Mismatches 522; Indels 533; Gaps 67;

QY 152 PILLHRRPFVRQPPADLRPDDEEVFYCKVTNEIFRHYDDFFERTILCNLSVMSCAVTGRPG 211
 Db 3 PLLGRKPPPLVKP--LPGEPLFTIPHTQEAFTREEYEARELRYSERIWTCKSTGSSQ 59

QY 212 LTYQALRESEKARQNL-QSPF---EPLIIPVLYLTSITHRSLRHEICDDIFAYVKDRYF 267
 Db 60 LTHKEAWEEBQVALLKEEPFAPWYKLVLEMHV---HNTASLEKLVDTAWLEIMTKYA 115

QY 268 VEETVEIRNNGARLOCTILEVLP-----PS-----HONGFANGHVN 304
 Db 116 VGEBCDFEVGKEMKLVKVIKHLPLEKVDDEATEKKSAGDPSKDKENSQIAQDHQK 175

QY 305 SVDGFTIISDDSETOS-----CSFQNGKKDAIDPLL-FKYKVQPTKKBELH 352
 Db 176 K---ETVVKEDGREGSINDRARRSPKPLPSLKKGERKWAPPKPLPKHYDV---KLQN 228

QY 353 ESAI---VKATQISRRKHLFRDKLKLFLKQHC-----EPOGVK---IKASLSLYK 400
 Db 229 EDKIISNPADSLIRTERPPNKEIVRYFIRHNALRAGTCENAPWVVEDELVKYKSLPS-- 286

QY 401 LAEQDFSYFFDDPTTFIFSPANRRG-----RPPKRIHSQEDN----- 440
 Db 287 ---KPSDFLLDPYKWTINLSTKTKNTGSPDRKPSKK---SKTDNSLSLPLNPKLWCH 339

QY 441 VANKOTLA-----SYRSKATKER----- 459
 Db 340 VHLKSLSGSLPKVKNKSNKSPKBEHLEMMKMSPNKLTNTHFIPKKGPPAKPKGKHS 399

QY 460 KLLK-----QEMKSLAFKAKLKR----- 479
 Db 400 KPLKAKGRSGKILNQKSTGNSKSPKGLKTPKTKMQWTLDDMAKGTQKMTAPRNSGG 459

QY 480 -----EKADALEAKKKEKEDKEKR----- 503
 Db 460 TPTRTSSKPHKHLPPAALHLIAYKENKDKREDKRSALSCVI SKTARLLSSEDRARLPEELR 519

QY 504 KIVEEE-----RLKKKEERKLVKEREKELRE--EKRKVYVLKOWSKPR- 549
 Db 520 SLVQKRYELLEHKKKSWASMSQORKEYLKKREELKKLKKKAKERREKEMLERLEKQR 579

QY 550 -EDMECDLKPPEPTPVKT--RLPPEITFGDALMVLEFLNAPFELQDFIDQDFPPGVTLFV 606
 Db 580 YEDQELTG-KNLPAPRLVDTPEGLNTLFGDVAMVVEFLSCYSGLLLPDAQYP--ITAVS 636

QY 607 LEEALVGNDSGPELCELLFFFLTAIFQAIABE--EEVAKQLTDADTKGCSLSKSLDLS 664
 Db 637 LMEAL-SADKGG-----FLYNRLVILLOTLLQDEIAB---DYGELGKMLSEIPLTL 685

QY 665 CTLSEILHLILASGADYTSANAKRYQKRGGFADTDACMELRLNSPLVKLSSTSVY 724
 Db 686 HVSVELVLCRRSDVQSESGSDT-----DNKDSAAEDNEVQDEFLEKLETSFF 738

QY 725 DLTFQKMKILHALCGKLLTLVSTRDFIEDYVDYLROAKQBFRELKABQHRKEREAAAR 784

Db 739 ELTSEKLOILTALCHRLIMTYSVQD-----HMETROOMSAB 775
 QY 785 IRKKEEKLQEQOKMEKQKLEKDE---ORNSTADISIGEEREEDPDTSTIESKDEOK 841
 Db 776 LWKERLAVLKEENDKRAEKQKREMEAKNKENGKVGELGKTDRK-----KEIV 825

QY 842 ELDDQMFTEDEDDPGSHKGRG---RGKRGQNGKFETTRQBOQINCVTRELTLADBEELAK 897
 Db 826 KFEPQVDTEADMISAVKSRRLAIQAK-----KEREIQE-----REMKVLERQA-E 872

QY 898 QEHORKEKELEK-IQSAIACNIF---PLGRDMYRYWIF-PSICGLFTEEDYSGLT 951
 Db 873 EERIRKXKAAAEKAPQEGIAKAKLVMRRTPTGTDRNHRYLWLFSDVEVGLFTIEK---GMV 929

QY 952 EDMLLPRPSSFQNNVQSDPOVSTKTGEP-----LMSESTSNIDQO-PRDHSVOL 1000
 Db 930 HDSI-----DYRFNHCHCKDHTVS---GDGYCPRSKKANLGKXNASMNTQGTATEVAVET 981

QY 1001 PKPVHKNRWCFYSCEQLDQLEIALNSRGHRESALKETLLQEKSRICAQLARFSE--EK 1058
 Db 982 TTPKQGNLWFLCDSQKELDELLNCLHPQIGRESQKLERL-EK-----RYQDIHS 1031

QY 1059 FHFSDKPOPSKPYTSRGRSSNAYDPQMCACQLELRLDFLLDIEDRIYQGTIGAKV 1118
 Db 1032 IHLARKPNLGLK-----SCDGNQ-----ELLNFRSLDLTEVATRLQKGLGYVEE 1076

QY 1119 TDRHWSALSGRYELLSEENKENGIIKTVNEDEVEMEIDBQTKVIVK--DRLLAGIKTE 1176
 Db 1077 T-----SEFA---RVISLEKL-----KQFGEVIALQASVKKFLOGFWAPKQK 1118

QY 1177 TPSTVSTNASTPQSVSVVHYLAMALFOIEOGIERFLKAPLDASDSGRSVKTVLDRWE 1236
 Db 1119 RRKLQSEDSAKTEEVE-----EKQWVEEAKVAS-----ALEKWT 1154

QY 1237 SLLSASLSQVFLHLSTLDRSVINWSKSLNARCKIKRKGDAENMVLCDGCDRGHTTVCV 1296
 Db 1155 AIREAQTFSRMHVLLGLMDACIKWDSAEANARCKVCRKKGEDDKLILCECNKAPHLFCL 1214

QY 1297 RPKLKTYPGEGWGFCEPCRKQRCRLSPRQPSLESDDSDMSGREDVDGDEEGQS 1356
 Db 1215 RPALEYPDGQWQCPACQATARNRGRNTYESASESDESDDESEEE---EEBEE 1270

QY 1357 EEEYEVEQDDEDDSQEEEEVSLPKRGPQVRL-PVKT-RGKLSSSFSSRGOQOEGRYPS 1414
 Db 1271 EEEYEV-----AGLRLRPKRTIRGKHSVTPPAARSGRRPGKXPH 1310

QY 1415 RSQOSTPKTTVSSKTGRSLRKINGAPPTETKSLRIASRSTRSHGHPLOADVVELLSPPR 1474
 Db 1311 STRSQPK-----APPVD-----DAEVDL----- 1330

QY 1475 KRRGRKSANNTPENSPNPFNFRVIATKSSSEOSRSVNIATSKLSQESSEKRRCKRQSP 1534
 Db 1331 -----VIQTKRSSRRQSL----- 1343

QY 1535 SPVTILGRRSSRGQGVHLSAFEOLVVELRHDDSWPFLKLVSKIQVDDYDIKKPIAL 1594
 Db 1344 -----ELQKCEELHKIVKYSWPEPVPTRDEADYDVIHTHPMF 1386

QY 1595 NIIEKNKCYKLASFPIDDIELMFNSCNFPYNPRT 1631
 Db 1387 QTVQNKSCGYSRVSQBFSLTDMKQVFTNAEYVNCGRS 1423

RESULT 15
 Q8UVR4
 ID Q8UVR4 PRELIMINARY; PRT; 1079 AA.
 AC Q8UVR4;
 DT 01-MAR-2002 (TremBLrel. 20, Created)
 DT 01-MAR-2002 (TremBLrel. 20, Last sequence update)
 DT 01-MAR-2004 (TremBLrel. 26, Last annotation update)
 DE Williams syndrome transcription factor (fragment).
 OS Xenopus laevis (African clawed frog).

CC -|- SIMILARITY: Contains 1 methyl-CpG-binding (MBD) domain.
 CC -|- SIMILARITY: Contains 1 PHD-type zinc finger.
 CC -|- CAUTION: Ref.5 sequence differs from that shown due to a
 CC frameshift in position 731.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----

DR EMBL; AB032255; BAA89212.1; --
 DR EMBL; AB040909; BAA96000.2; ALT INIT.
 DR EMBL; AL834381; CAD39044.1; ALT INIT.
 DR EMBL; BC012576; AAH12576.1; ALT FRAME.
 DR EMBL; AK027612; BAB55231.1; ALT_FRAME.
 DR EMBL; AL080173; CAB45759.1; --
 DR F1R; T12495; T12495.
 DR HSP; Q9UIG0; 1F62.
 DR Genew; HGNC:963; BAZ2B.
 DR MIM; 605683; --
 DR InterPro; IPR001487; Bromodomain.
 DR InterPro; IPR004022; DDT.
 DR InterPro; IPR001739; Methyl-CpG_bind.
 DR InterPro; IPR001965; Znf_PHD.
 DR Pfam; PF00439; Bromodomain; 1.
 DR Pfam; PF02791; DDT; 1.
 DR Pfam; PF01429; MED; 1.
 DR Pfam; PF00628; PHD; 1.
 DR PRINTS; PR00503; BROMODOMAIN.
 DR PROSITE; PS00633; BROMODOMAIN_1; FALSE_NEG.
 DR PROSITE; PS50014; BROMODOMAIN_2; 1.
 DR PROSITE; PS50827; DDT; 1.
 DR PROSITE; PS50982; MED; 1.
 DR PROSITE; PS01359; ZF_PHD_1; FALSE_NEG.
 DR PROSITE; PS50016; ZF_PHD_2; 1.
 KW Alternative splicing; Bromodomain; Coiled coil; DNA-binding;
 KW Nuclear protein; Transcription regulation; Zinc-finger.
 FT DOMAIN 543 614
 FT DOMAIN 891 956
 FT ZN_FING 1735 1785
 FT DOMAIN 1881 1951
 FT DOMAIN 45 69
 FT DOMAIN 73 79
 FT DOMAIN 399 470
 FT DOMAIN 641 672
 FT DOMAIN 687 865
 FT DOMAIN 706 875
 FT DOMAIN 1100 1143
 FT DOMAIN 1138 1179
 FT VARSPPLIC 437 534
 FT VARSPPLIC 593 626
 FT CONFLICT 137 137
 FT CONFLICT 226 226
 FT CONFLICT 627 627
 FT CONFLICT 722 722
 FT CONFLICT 799 799
 FT CONFLICT 1128 1128
 FT CONFLICT 1183 1183
 FT CONFLICT 1195 1195
 FT CONFLICT 1453 1453
 FT CONFLICT 1838 1838
 FT SEQUENCE 1972 AA; 220709 MW; D4F105C927B91569 CRC64;
 Query Match 8.5%; Score 731.5; DB 1; Length 1972;
 Best Local Similarity 21.4%; Pred. No. 6.6e-22;
 Matches 335; Conservative 196; Mismatches 424; Indels 613; Gaps 53;

QY 420 SPANRRGRPP-----KRIHSQEDNVA--- 442

Db 664 SMRRRKGRPPNVGNAEFLDNADAKLLRKLQAEIARAAQAIKLRLKQKQEQARVAKSA 723
 QY 443 --NKQTLASYRSKATKRDKLLKQBE---MKSFAE-----KAKLKREKADA- 484
 Db 724 KQQAIMAAEKKRQKEQIKINKQKEIKRIQIRMEKELRAQQLLEAKKKKEEANAAN 783
 QY 485 -LEAKK--KEKE-----DKEKKREBKIVBEERLKKKEERLKVVERKE-- 527
 Db 784 LLEAEKRIKEMRRQAVLLKHQERRRRQHMLMKAMEARKKAEKELRQKREKDEKR 843
 QY 528 --REKLREKRYVEYLQWSKPREDMCEDDLKELPE-PTPVKTRLPPIPIFGDALMVLSE 584
 Db 844 LNKERKLEQRLEMAKELKKPNEDMCLADQKPLPELPFGLVLSGTSFSDCLMVVQF 903
 QY 585 LNAFGEF--FDLQDEFFPGCVTLVELEAEALVG-NDSEGLCELLFFLTAFIQAIAEES 641
 Db 904 LENFKVLGFDVNIQVFN---LSVLQEGLLNIGDSMGVQDILLVLLSA---AVCDPGLI 957
 QY 642 VAKEQLTDADTKGCSLKSLLDSCCTLSEILRHILASGADVTSANAKYRKRGGPDATD 701
 Db 958 TGYKAKT---ALGEHLLNVGVNRDENVSEILQIF----- 987
 QY 702 DACMELRLSNPSLVKKLSSTSVYDLTPGKKMILHALCGKLL---TLVSTRDFIEDYVDI 758
 Db 988 ---MEAHCGQTELTESLTKTAPQAHTPAKASVLAFLINELACSKSVSEIDKNIDYMS 1044
 QY 759 LRQAKQEFRELKAEQHRKEREAAARIRKKEEKLKEQEKMEKQEKLEKEQENSTAD 818
 Db 1045 LRDRKVV-----EGKLR-----KLRIHAHKTGKRTSGG 1075
 QY 819 ISIGEE-----REDFTSIESKDTQEKELDQDMFTDEDDPGSHKRRGK 865
 Db 1076 IDLGEQHPGLGTPGKRKRKGGSDYDDDDDDDDGDEDEDEDEDEDEDEDEDEDEDE 1131
 QY 866 RGQGFKEFTQEQINCVTRELLTADSEE---ALKQEHQKKEKLEKIQS-----A 914
 Db 1132 KTD-----ICDEDEGDAQASVEELEKQIEKLSKQSQSYRRKLQFA 1172
 QY 915 IACTNIFPLGRDRMYRWIPPSIPGLFIE--EDYSGLTEDMLPRPSSFQNNVQSDPQ 972
 Db 1173 SHLSRVWFGQDRYRRYWIIPQCGQIFVEGMESEGLEEIIAKEREKLLKASVQIKEM 1232
 QY 973 VSTKTGPELMSESTSNIDQG-----PRDHSVQLP 1001
 Db 1233 FET--SGDSLNCSTNTHCQKEDLKEKNTNLFQKPGSFKLSKLEVAKMPPESEVMTP 1291
 QY 1002 KP-----VHK-----PNRWCFYSSCEQ 1018
 Db 1292 KPNAGANGCTLSYQNSGRKSLGSVQSTATQSNVEKADSNNTFGTSSGPGK--FYSPLPN 1349
 QY 1019 LDOLLEALNSGRHRESAL-----KETLLQEKSRICAQALARESEKHFSDKPPD--- 1068
 Db 1350 -DOLLKTLTEKNRQWFSLLPRTPCDDTLTHADMSTASLVTPQSPPPSKSPPTAPLGS 1408
 QY 1069 -----SKPTYSRG-----RS 1078
 Db 1409 SAQNPVGLNPFALSPQVKGVSMMGLQFCGWPTGVVTSNIPTLTVSLGSLGSLGSEGN 1468
 QY 1079 SNAYDPSQWCAEQ-----LEL-RLRDF-----LLDI 1104
 Db 1469 GNSFLTSNVASSKSESPVQNEKATSAQPAAEVAVKPVDFPSPKPIPEMQFGWRIIDP 1528
 QY 1105 EDRIYQGTGLAI-----KVTDRH---IWRSALESGRYELLS--EENKENG 1145
 Db 1529 ED-----LKALLKVLHURGIREKALQIQIKHLDYITQACUKNKDVAIIELNENE 1582
 QY 1146 IKTV--NEDVEEMIDEQKIVIKDRLLGKITETPTSTVSTNASTPQSVS----- 1192
 Db 1583 TRDIVNWSVEQAMEMDLSVLQVEDLERRVASASLQVKGWMCPEPASEREDLVYFEHK 1642
 QY 1193 -----SVVHYL-----AMALFOIEQGIERRFLK--APLDAS 1221

```

Db 1643 SFTKCKEHGDEFTGEDESSAHALERKSDNPLDIAVTRLADLERNRIERBEDIAP----- 1698
Oy 1222 DSGRSYKTVLDWRRESLSSASLSOVFLHSLTLDRESVLSKSIILNARCKIKRKGDAENM 1281
Db 1699 -----GLRWRRALSEARSAAQVLCIOOLQKSIAMKSVKVCQICRKGDEBELL 1750
Oy 1282 VLDGCGDRGHHTYCVRPKLVPEGDWFCPECRPKQ-----RCRLSPRQPSLESDEVDV 1336
Db 1751 LLCDCGDKGCHTYCHRPKITIPDGDWFCPACIAKASQOTIKIKLHVKGKKTWESKK-- 1808
Oy 1337 EDSMGDEVDGDEEQSQSEEEVEYEQDEDDSQEEBEVSLPKRGRPQVRLPVKTRGKL 1396
Db 1809 -----GKKVLTGDTGTE-----DED----- 1822
Oy 1397 SSSPSRQQQEPGRYPSPRSQSQSTPKTVTSKTSGRSLRKINSAPPTETKSURIASRSTRH 1456
Db 1823 -----SASTSSSL----- 1830
Oy 1457 SHGPLQADVVELLSPRKRGRKRSANNTPENSFPNFRVIATKSSPSQSRSVNIASKLS 1516
Db 1831 -----KGNKD-----LKKRKEVNTSIN-----LS 1851
Oy 1517 LQESKRCRKRKROSPSPVTLGRSSRGQGVHLSAFQLVVLVVRHDDSWPFLKLV 1576
Db 1852 KQESFTSVKPKRDDSK-----DLALCSMILTEMETHEDAWPFLLPV 1893
Oy 1577 SKIQVPDYDIKKPIALNIIRKVNKCEYKLASEFDDIELMFSNCPYVNPRTNSEKA 1636
Db 1894 NLKLVPGYKVIKFMDFSTIREKLSQGYNPLETALDVLRLVDFNCFTFEDSDIGRA 1953
Oy 1637 GTRLQARF 1644
Db 1954 GHNWKYF 1961

RESULT 18
BAZA_HUMAN
ID_BAZA_HUMAN STANDARD; PRT; 1878 AA.
AC Q9U1F9; O00536; O15030; Q96H26;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Bromodomain adjacent to zinc finger domain 2A (Transcription
DE termination factor-I interacting protein 5) (TTF-I interacting protein
DE 5) (Ttp5) (hWALp3).
GN Name=BAZ2A; Synonyms=K1AA0314, TIP5;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RX MEDLINE=20130112; PubMed=1062543; DOI=10.1006/geno.1999.6071;
RA Jones M.H., Hamana N., Nezu J.-I., Shimane M.;
RT "A novel family of bromodomain genes.";
RL Genomics 63:40-45 (2000).
RN [2]
RP SEQUENCE OF 332-738 FROM N.A.
RC TISSUE=Lung;
RA Jansa P., Grummt I.;
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 639-1878 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=97349984; PubMed=9205841;
RA Nagase T., Ichikawa K.-I., Nakajima D., Ohira M., Seki N.,
RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. VII.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro.";
RL DNA Res. 4:141-150 (1997).
RN [4]

```

```

RP SEQUENCE OF 1038-1878 FROM N.A.
RC TISSUE=Lymph;
RX MEDLINE=4238257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Schetz T.E.,
RA Brownstein M.J., Uedin T.B., Toshiyuki S., Carninci P., Frange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butcherfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,
RA Smerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RP Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
CC -!- FUNCTION: May play a role in transcriptional regulation
CC interacting with ISWI. May serve a specific role in maintaining or
CC altering the chromatin structure of the rDNA locus (By
CC similarity).
CC -!- SUBUNIT: Together with ISWI/SNF2h, it forms a complex termed NORC
CC (nucleolar remodeling complex).
CC -!- SUBCELLULAR LOCATION: Nuclear. Colocalizes with the basal RNA
CC polymerase I transcription factor UBF in the nucleolus.
CC -!- TISSUE SPECIFICITY: Expressed at moderate levels in most tissues
CC analyzed, including heart, brain, placenta, lung, skeletal muscle,
CC kidney and pancreas.
CC -!- SIMILARITY: Belongs to the WAL family.
CC -!- SIMILARITY: Contains 4 A.T hook DNA-binding repeats.
CC -!- SIMILARITY: Contains 1 bromodomain.
CC -!- SIMILARITY: Contains 1 DDT domain.
CC -!- SIMILARITY: Contains 1 methyl-CpG-binding (MBD) domain.
CC -!- SIMILARITY: Contains 1 PHD-type zinc finger.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB032254; BAA89211.1; -
DR EMBL; AF000422; AAB60864.1; -
DR EMBL; AB002312; BAA20773.1; -
DR EMBL; BC008965; AAB08965.2; -
DR HSSP; Q9UIG0; 1F62.
DR Genew; HGNC:962; BAZ2A.
DR MIN; 605682; -.
DR GO; GO:0005731; C:nucleolus organizer complex; NAS.
DR GO; GO:0003677; F:DNA binding; NAS.
DR GO; GO:0030528; F:transcription regulator activity; NAS.
DR GO; GO:0006338; P:chromatin remodeling; NAS.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; NAS.
DR InterPro; IPR000637; A-T_hook.
DR InterPro; IPR001487; Bromodomain.
DR InterPro; IPR004022; DDT.
DR InterPro; IPR001739; Methyl-CpG_bind.
DR InterPro; IPR001965; Znf_PHD.
DR Pfam; PF02178; AT_hook; 4.
DR Pfam; PF00439; Bromodomain; 1.
DR Pfam; PF02791; DDT; 1.
DR Pfam; PF01429; MBD; 1.
DR Pfam; PF00628; PHD; 1.
DR PRINTS; PR00503; BROMODOMAIN.
DR SMART; SM00384; AT_hook; 4.

```

DR SMART; SM00297; BROMO; 1.
 DR SMART; SM00571; DDT; 1.
 DR SMART; SM00391; MBD; 1.
 DR SMART; SM00249; PHD; 1.
 DR PROSITE; PS00633; BROMODOMAIN_1; FALSE_NEG.
 DR PROSITE; PS0014; BROMODOMAIN_2; 1.
 DR PROSITE; PS50827; DDT; 1.
 DR PROSITE; PS50982; MBD; 1.
 DR PROSITE; PS01359; ZF_PHD_1; FALSE_NEG.
 DR PROSITE; PS0016; ZF_PHD_2; 1.
 KW Bromodomain; Coiled Coil; DNA-binding; Nuclear protein; Repeat;
 Transcription regulation; Zinc-finger.
 FT DOMAIN 519 590 MBD.
 FT DNA_BIND 622 634 A.T hook 1.
 FT DNA_BIND 643 655 A.T hook 2.
 FT DOMAIN 821 886 DDT.
 FT DNA_BIND 1159 1171 A.T hook 3.
 FT DNA_BIND 1377 1389 A.T hook 4.
 FT ZN_FING 1649 1699 PHD-type.
 FT DOMAIN 1783 1853 Bromodomain.
 FT DOMAIN 633 772 Lys-rich.
 FT DOMAIN 666 765 Coiled coil (Potential).
 FT DOMAIN 1185 1250 Glu-rich.
 FT DOMAIN 1263 1384 Pro-rich.
 FT DOMAIN 1732 1735 Poly-Arg.
 FT CONFLICT 574 574 L -> V (in Ref. 2).
 FT CONFLICT 700 700 L -> Q (in Ref. 2 and 3).
 FT CONFLICT 720 720 H -> Q (in Ref. 2 and 3).
 FT CONFLICT 727 738 SKAEKEKGTQ -> KIKKKKKKKK (in Ref. 2).
 FT CONFLICT 785 785 K -> R (in Ref. 3).
 FT CONFLICT 951 951 P -> L (in Ref. 3).
 FT CONFLICT 1005 1006 GR -> EG (in Ref. 3).
 FT CONFLICT 1163 1163 G -> S (in Ref. 1).
 FT CONFLICT 1166 1166 R -> L (in Ref. 1).
 FT CONFLICT 1172 1172 S -> F (in Ref. 1).
 FT CONFLICT 1178 1178 L -> F (in Ref. 1).
 FT CONFLICT 1202 1202 A -> V (in Ref. 1).
 FT CONFLICT 1292 1292 P -> L (in Ref. 1).
 FT CONFLICT 1295 1295 L -> F (in Ref. 1).
 FT CONFLICT 1313 1313 P -> L (in Ref. 1).
 FT CONFLICT 1407 1410 Missing (in Ref. 4).
 FT CONFLICT 1416 1416 R -> P (in Ref. 1).
 FT CONFLICT 1541 1541 R -> P (in Ref. 1).
 FT CONFLICT 1571 1571 E -> K (in Ref. 1).
 FT CONFLICT 1616 1616 V -> I (in Ref. 1).
 FT CONFLICT 1622 1622 E -> Q (in Ref. 1).
 FT CONFLICT 1629 1629 Q -> H (in Ref. 1).
 FT CONFLICT 1636 1636 Q -> H (in Ref. 1).
 FT CONFLICT 1739 1739 R -> K (in Ref. 1).
 FT CONFLICT 1754 1754 G -> R (in Ref. 1).
 SQ SEQUENCE 1878 AA; 208480 MW; 400970CA68234317 CRC64;

Query Match 8.1%; Score 695.5; DB 1; Length 1878;
 Best Local Similarity 20.2%; Pred. No. 1.9e-20;
 Matches 312; Conservative 195; Mismatches 392; Indels 643; Gaps 47;

QY 424 RRRGRPPK-RIH-----ISQSDN-----VANKOTLASYSKATKRDKLLKQEMKSL 470
 DB 644 RGRGPPVKITELLNKTNDNPLKLEAQETLNEEDXAKIAKSKKMKQKVRGECITTI 703
 QY 471 AFEKAKLKREKADAL---EAKKEKEDKEK---KREELKKIVEE-----BLKKEE--- 516
 DB 704 QQARNKKEQETKSLKHKEAKKSKAEKKGTKQKLEKVKREKKEKVKMKKEEVTK 763
 QY 517 -KERLKVREKEREKLEKREKRYEYLKQWKPREDMCDLDELPEPTPVK-TLPPPI 574
 DB 764 AKPACKADKTLATQRRLEERQKQMILEMKKPTEDMCLTDHQPLPDRFVRPGLTLPSPA 823
 QY 575 FGDAIMLVLEFLNAFGL--FDLQDFPDGVTVLEVEALV-GNDSEGPLCELLPFFLTAI 631
 DB 824 FSDCLTIVEFLHSFGKVLGFPDAKVP---SLGVLOEGLLCQGSLSGEVDLLVRLKAA 880
 QY 632 F-----QAIAEEEEVAKEQLTDADTKGCSLKSLDLSCTLSLRLHLILASGADV 682

881 LHDPGFPSYCQSLKILGEKVSEIPLTRDN-----VSEILRCLFLMAYGE- 924
 683 TSANAKYRYQKRGDFDADDACMELRLSNPSLVKKLSTSVYDLTPGKMKIL-----HAL 738
 925 -----PALCDRLRTQPFQAQPPQQAQAAVAPFVHEL 955
 739 CGKLLTLVSTRDFIEDYVILRQAKQFRELKASQHRKEREEAAARIKRKEEKLKEOEQ 798
 956 NGSTL-----IINEIDKTLESMS--YRKNKWIIVEGLRLKTLAK----- 995
 799 KMKEKQKLEKDEQORNSTADISIGEEREDFDTSIESKDTQEKELDQDMFTEDDDPSH 858
 996 -----RTGRSEVEMGRPE-----ECIGRRRRSSIMBETSMESEESIAA 1037
 859 KRGRGRKRG-----QNGFKEFTREQEINCVTRELLTADEEALKQEHQKELLEKIQ 912
 1038 VPGRRGRDGEVDATASSIPELERQ-----LEKLSKROLFRKLLHSSQ 1082
 913 SAIACINIFPLGRDMYRRYWFPSIPGLFIEEDYSGLTEDMLLP----- 957
 1083 MLRAVS-----LGQDRYRRRYVWLPYLAGIFVEG-----TEGNLVPEEVIKKETDSLKVA 1133
 958 -----RPSSFQ-----NNVQSQD----- 970
 1134 HASLNPALFSMKMELAGSNNTTASSPARARGRPRTKPGSMQPRHLKSPVQGDSEPOQA 1193
 971 -----PQV-----STKTGE 979
 1194 LOPEAQLHAPQOPQLOLQSHKGFLQEGSPSLGQSOHDLSSQSFLSWLSQTSHS 1253
 980 PLMSESTSNIDQGR-----DHSVQLP----- 1001
 1254 SLSSSVLTDPSSPGKLDPAQPPPEPDEASSDPQALWFENISAQMPCNAAPTPPP 1313
 1002 -----KPVHKPNR----- 1009
 1314 AVSEDOPTSPQQLASSKPMNRPSAANPCSPVQFSSTPLAGLAPKRRAGDPCGEMPQSTG 1373
 1010 -----WCFYSSCEOLDQDLEIALNSRGH 1031
 1374 LGQPKRRGRPPSKFFKQWEOBYLTQTAQVPPPMCSGWWIRDPMDLMDALKALHPGI 1433
 1032 RESALKETLLQEKSRICAQLARFSEKHFSDKQPDSPKPTYSRGRSSNAYDPPQCAEK 1091
 1434 REKALHKLNRDFLQEVCLRPSADPIF-----EPQLPAFQEGIMS--NSPKETVE- 1485
 1092 QLELRDRDFLDIEDRIYQGTGLGAIKVTDRIHWSALESGRYELLSEENKENGIIKTVNE 1151
 1486 -TDLAVLQWVEELRQVIM-----SDLQIRGWTCPSPD-----STRE 1521
 1152 DVEEME--IDBQTKVIVKDRLLGIKTETPTSTVSTNASTPQSVSSVWHYALMALFOIEGI 1209
 1522 DLAYCEHLSQSQEDITWRGR--GREGLAPQRTKNP-----LDLAVMLAALEQNV 1570
 1210 ERREKAPLDASDSGRSYKTVLDR-----WRESLLSSA 1242
 1571 ERRYLREPL-----WPTHEVLEKALLSTPNGAPEGTTETSEYITPRIVRQTLERCR 1625
 1243 SLSQVFLHSLTDRSVTWSKILNARCKI CRKKGDAENMLVCDGCDRGHHTYCVRPKUKT 1302
 1626 SAAQVCLCLGQLERSIAWEKSVNKVTCLVCRKGDNDLFLLLCDGCDRGCHTYCHRPKMEA 1685
 1303 VPEDGMCPCECPKQRCRRLSFRQPSLESDEDDVDSNGSGDDEVDGDEEGSEEEYE 1362
 1686 VPEGDWFTCTVLAQQ-----VEGEFTQKPGPKGQKRRKSGYS 1723
 1363 VEQDEDSQEBEEVSLPKRGRPVRLPVKTRGKLSSSFSSRGOQEPGRYPSRSQOSTPK 1422
 1724 LNFSEGD-----GR---RRVLLRGR-----ESPAAGPRYSEEG--- 1754
 1423 TTVSSKTSRSLRKINSAPPTETKSLRIASRSTRSHGFLQADVVELLSPPRRKGRKSA 1482

Db	1755	-----LSPSKRRR-----	1762	Db	791	AKPACKADTKLATQRLEERQORQOMILEEMKPTEDMCLTDHQPDPDFSRVPLTSPGA	850
Qy	1483	NNTPENSFNFVRVIATKSSEQSRVNIASKLSLOESESRCRCRKQSPFPTLGR	1542	Qy	575	FGDALMVLLELNARFEL--FDLODEPPDGVTLVLEBALV-GNDSEGPLCELLFFFLITAI	631
Db	1763	-----LSMRNHS-----	1770	Db	851	FSDCLTIVEFLHSGVLGFDPAKVP---SLGVLOEGLLCCQGSLSGEVQDLLVRLLKAA	907
Qy	1543	SSGRQGVHELSAFEQVLVLRHDDSPFFKLVSQIOPDYDIKKPIALNIIREKVN	1602	Qy	632	F-----QATAEEEEEVAKELQTDADTKGCSLSKSLDLDSTLSEILRLHILHAGADV	682
Db	1771	-----DLTFCIILMEMESHDAWPFLEPNVRLVSGYRIIKNPMDFSTWRRELL	1821	Db	908	LHDPFPFSYQCSLKILGEKVSEIPLTDN-----VSEILACFLMAYGVE-	951
Qy	1603	KCEYKLASEFDDIELMFSNCFEYNPNRNTSBAKAGTRLOAF	1644	Qy	683	TSANAKYRQKRGFDATDACCMELRSLNSPLVKLSSTSVYDLTPGEKWKIL-----HAL	738
Db	1822	RGYTSBEFAADALLVFNQCNEEDSEVGKAGHIMRRFF	1863	Db	952	-----PALCDRLATQPPQAQPOPOCKAAVLAFLVHEL	982
RESULT 19							
Q68DI8	ID	Q68DI8	PRELIMINARY; PRT; 1905 AA.	Qy	739	CGKLLTLVSTRDFIEDYVDILROAKQEFRELK-AEQHKKERBEAAARIKKREKLEKQ	796
AC	Q68DI8;	25-OCT-2004 (TremBLrel. 28, Created)		Db	983	NGSTL-----IINSIDKTESMSYRNKNKIVGRLRLRLKTVLAKRTGRSEVEBGP	1034
DT	25-OCT-2004 (TremBLrel. 28, Last sequence update)			Qy	797	EQKM-KEKQEKLEDEORNSTADISIGEEREEDFDTSIESKDTEQEKELDQDMFTEDDDP	855
DT	25-OCT-2004 (TremBLrel. 28, Last annotation update)			Db	1035	EECLGRRSSRIMEE-----TSGMEE-----EEEES	1061
DE	Hypothetical protein DKFZp781B109.			Qy	856	GSHKRGRRKRG-----QNGFKFETRQEQINCVTRELLTADDEEALKQSHQRKEKELLE	909
GN	Name=DKFZp781B109;			Db	1062	IAAVPGRGRGRDGEVDATASSIPELERQ-----IEKLSKRLPFRKKLLH	1106
OS	Hom sapiens (Human).			Qy	910	KIQSAINACTNIFPLGRDRMYRYWIPPSIPGLFIEDYSGLTEDMLP-----	957
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			Db	1107	SSQMLRAVS-----LGQDRYRRYVWLPYLAGI FVEG-----TEGNLVEEVIVIKETDSLK	1157
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			Qy	958	-----RPSSFO-----NNVOSQD-----	970
OX	NCBI_TaxID=9606;			Db	1158	VAAHASINPALFSMKMELAGSNTTASSPARARGPRKTKFGSMQPHLKSPVRQDSEQP	1217
RN	SEQUENCE FROM N.A.			Qy	971	-----PQV-----STK	976
RG	TISSUE=Fetal kidney;			Db	1218	QAQLQPEALHAPAQPOPQLQLQSHKGFEQEGSPLSLGQSHLSQSASFSLWLNQTQ	1277
RA	The German cDNA Consortium;			Qy	977	TGEPLMSESTSNIDQGR-----DHSVQLP-----	1001
RA	Bloecker H., Boecker M., Brandt P., Mewes H.W., Weil B., Amid C.,			Db	1278	SHSLLSSSVLTPDSSFGKLDPAPOPPEPEPEDEAESSPPQALWFNISAQPCNAAPT	1337
RL	Osanger A., Fobo G., Han M., Wiemann S.;			Qy	1002	-----KPVHKPNR-----	1009
RL	Submitted (AUG-2004) to the EMBL/GenBank/DBJ databases.			Db	1338	PPPAVSEDQTPSPQQLASSKPMNRPSAANPCSPVQFSSTPLAGLAPKRAGDPGEMPQS	1397
DR	EMBL; CR749379; CAH18232.1; -			Qy	1010	-----WCFYSCEOLDQIEALNS	1028
DR	InterPro; IPR000637; A+T hook.			Db	1398	PTGLGQPKRRGRPPSKFFKQMEQRYLTQLTAQPVPPPMCSGWWIIRDPEDLAMLKALHP	1457
DR	InterPro; IPR001487; Bromodomain.			Qy	1029	RGHRESALKETILQEKSRICAQLARFSEEFHFDKPPQPSKPTYSGRSNADVPQMC	1088
DR	InterPro; IPR004022; DDT.			Db	1458	RGIREKALHKLHNRDFLOEVCRLRPSADPIF-----EPQLPAFQEGINS--WSPKEKT	1510
DR	InterPro; IPR011011; FIVE_PHD_ZnF.			Qy	1089	AEKQLELRDRDRLDIEDRIYQGTGLGAIKVTDRHWRSALESGRYELLSSSENKENGIIKT	1148
DR	InterPro; IPR001739; Methyl-CpG_bind.			Db	1511	YE--TDLAVLQWVEELEQRVIM-----SDLIQRWTCPSPD-----S	1545
DR	InterPro; IPR001965; ZnF PHD.			Qy	1149	VNEDVEEME--IDEQTKVIVKRLLLGIKTPTSTVSTNASTPOSVSVVHYLAMALFOIE	1206
DR	Pfam; PF02178; AT hook; 4.			Db	1546	TREDLAYCEHLSDSQEDITWRGR--GREGLAPQRKTTNP-----LDLAVMLAALE	1594
DR	Pfam; PF02791; DDT; 1.			Qy	1207	QGIERRFLKAPLADSDGRSYKTVLDR-----WRSILL	1239
DR	Pfam; PF00628; PHD; 1.			Db	1595	QNVERRYREPL-----WPTHVEVLEKALLSTPNAGEGTTTTSYEITPRIRVWROTLE	1649
DR	Pfam; PF01429; PHD; 1.			Qy	1240	SSASLSQVFLHSLTDRSVIWSKIINARCKI CRKKGDAENMVLCDCCDGHHTYCVRPK	1299
DR	PRINTS; PR00503; BROMODOMAIN.			Db	1650	RCSAAQVCLICLQGLERSIAWESVNVKTVCLVCRKGNDFELLCDCCDGHCHYCRPK	1709
DR	SMART; SM00384; AT hook; 4.			Qy	1300	LKTVPEGDWFCPCRPKQRCRLSFRQRPSSLEDEVDVSMGGEDDDVDGDEEGSEEE	1359
DR	SMART; SM00297; BRÖMO; 1.			Db	1710	MEAVPEGDWFCVCLAQO-----VEGEFTQKPGFPKRGQRKS	1747
DR	SMART; SM00571; DDT; 1.						
DR	SMART; SM00391; MBD; 1.						
DR	SMART; SM00249; PHD; 1.						
DR	PROSITE; PS50014; BROMODOMAIN_2; 1.						
DR	PROSITE; PS50827; DDT; 1.						
DR	PROSITE; PS50016; ZF_PHD_2; 1.						
KW	Hypothetical protein.						
SQ	SEQUENCE 1905 AA; 211195 MW; F7C781990180A2E1 CRC64;						
Query Match							
Best Local Similarity 20.3%; Pred. No. 3.4e-20;							
Matches 314; Conservative 195; Mismatches 387; Indels 649; Gaps 48;							
Qy	424	RRGRGPPK-RH--ISOEDN-----VANKOTLASYSKATYKRDKLLK---QEEM	467	Qy	424	RRGRGPPK-RH--ISOEDN-----VANKOTLASYSKATYKRDKLLK---QEEM	467
Db	671	RGRGRPPVKITELLNLRPLKLEAQETLNEDKAKIAKSKKMKRQKVQEGECQTTI	730	Db	671	RGRGRPPVKITELLNLRPLKLEAQETLNEDKAKIAKSKKMKRQKVQEGECQTTI	730
Qy	468	KSLAFKAKLREKADALEAKKEKEDKEK---KREELKKIVVEE-----RLKKKEE---	516	Qy	468	KSLAFKAKLREKADALEAKKEKEDKEK---KREELKKIVVEE-----RLKKKEE---	516
Db	731	QGQARNKQKQTKLQKEAKKKAKEKQKTKQEKLEKVKREKKEKVKQKKEEVTK	790	Db	731	QGQARNKQKQTKLQKEAKKKAKEKQKTKQEKLEKVKREKKEKVKQKKEEVTK	790
Qy	517	-KERLKVREKERELREKRYVEYLKQWQKPREDMCEDDLKELPPTPVK--TRLPPEI	574	Qy	517	-KERLKVREKERELREKRYVEYLKQWQKPREDMCEDDLKELPPTPVK--TRLPPEI	574


```

QY 1360 EYEVEQEDDSQEBEVSPLPKGRQVRLPVKTRKGLSSFSRQOQEPGRYPRSQOS 1419
Db 1748 GYSLNFSEGD-----GR---RRVLLRGR-----ESPAGPRYSBEG 1781
QY 1420 TEKTTVSSKTRSLRKINSAPPTETKSLRIASRSTRHSHGPLQADVFFVLLSPRKRGR 1479
Db 1782 -----LSPSKRRR-- 1789
QY 1480 KSANTPENSFPNFRVIATKSSQSRSVNTASKLSLQESSKRCRKRQSPSPVTL 1539
Db 1790 -----LSMENHHS----- 1797
QY 1540 GRRSGRQGVHLSAFELQVLVLRHDSWFLKLVSKIQVDPDYDIKPIALNIIRE 1599
Db 1798 -----DLTFCFELMEMSHDAAWPFLPVPNRLVSGYRRIRKKNPDMFSTWRE 1845
QY 1600 KVNKEYKLASFIDDIELMPNCFEYNPNTSEAKAGTRLOAFF 1644
Db 1846 RLLRGYTSSEFAADALLVFNDQCFTNEDDSEVGKAGHIMRRFF 1890

RESULT 20
Q6Y194
ID Q6Y194 PRELIMINARY; PRT; 1698 AA.
AC Q6Y194;
DT 05-JUL-2004 (TremBLrel. 27, Created)
DT 05-JUL-2004 (TremBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TremBLrel. 27, Last annotation update)
DE Putative chromatin remodelling factor BAZ2A.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RA Ruzov A.; Meehan R.;
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY145834; AAN61105.1; -.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0006355; F:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR000637; A+T hook.
DR InterPro; IPR001487; Bromodomain.
DR InterPro; IPR004022; DDT.
DR InterPro; IPR001739; Methyl-CpG_bind.
DR InterPro; IPR001965; Znf_PHD.
DR Pfam; PF02178; AT_hook; 3.
DR Pfam; PF00439; Bromodomain; 1.
DR Pfam; PF02791; DDT; 1.
DR Pfam; PF01429; MBD; 1.
DR Pfam; PF00628; PHD; 1.
DR PRINTS; PR00503; BROMODOMAIN.
DR SMART; SM00384; AT_hook; 3.
DR SMART; SM00297; BROMO; 1.
DR SMART; SM00571; DDT; 1.
DR SMART; SM00391; MBD; 1.
DR SMART; SM00249; PHD; 1.
DR PROSITE; PS50014; BROMODOMAIN_2; 1.
DR PROSITE; PS50827; DDT; 1.
DR PROSITE; PS50016; ZF_PHD_2; 1.
SQ SEQUENCE 1698 AA; 192135 MW; F62D613AD9611B6F CRC64;

Query Match
Best Local Similarity 7.8%; Score 673; DB 2; Length 1698;
Matches 318; Conservative 180; Mismatches 413; Indels 544; Gaps 53;

QY 422 ANRRGRPPKRIHSQEDNVANKQTLASYSKATK--ERDKLLKQEMKSLAFKAKLKR 479
Db 540 AKRGRGPPK-----VKMDLLSKADAKLLRKLKNQDILSDSEKVCCKLKKWRR 590
QY 480 EKAD---ALEAKKKEKE--DKKKREELKIVVEERLKKKEKRLKVEREK----- 526

```

```

Db 591 KARNQEALEAKKLUKEIKKEEKKQKTKAKNQSKAKNQEKTRRQPKQKAPVVKQPD 650
QY 527 -----EREKLEEREKYYEVLKQWSPKREDMECDLKEPPE-PTPVKTRLPPEIFGDALM 580
Db 631 RKQLAQORRLKEREKQOF-ILEELKKPTEDMCLPDHQQLPDPFPCVPGFLPSCASDCULT 709
QY 581 VLEFLNARFELFDLODEFPDGVTLLEBEAL--VGNDSGELCELLFFPLTAI----- 631
Db 710 TVEFLHSYGVKVLGL-DEAKDIPSLCTLOEGLPKVG-DSLGEVQDILLKLLQAAMINPGLP 767
QY 632 -----FOAIAEBEEVAKELTDADTKGCSLKLSDLSCTLSIILRLHILASGADVTSAN 686
Db 768 PYCOSLKITLGBKVSIV-----LSRENVSEPRJIFLEAYGGDI----- 805
QY 687 AKRYQKRGGFDDTADACMELRSLNPSLVKKLSSTSVYDLTPGEKWKILHALCGKLLTLV 746
Db 806 -----QLCDSLRL-SHPPQAH-----APHIKTALIAFVNEL----- 835
QY 747 STROPFIEDYVILRQAKQEFRELKABQHRKEREAAAIRKRKEKLEKQEQMKKEQK 806
Db 836 NASSLIISEIDKTLENMSHYR-----KHKWIIEGKIRLKFALSKRKSSESQITTEVS 890
QY 807 LKEDEQRNSTADISIGBEEREDFOTSIESKDTEQKELQDQDMEFTEDEDDPGSHKRGK 866
Db 891 LRRSERNA-----EENDELSIDES-----AIQKDYVQEEVDIPPS----- 927
QY 867 GONGKFEPTROEQINCVTRELLTADDEEALKQEHQKELKEIKQSAIACINIFPLGRD 926
Db 928 --TSVVELERO-----IEKLTROMFFRKILGSSORLRTVC-----LQY 966
QY 927 RMYRYNIFPSPGLFIE-----EDYSGLTEDM-----LLPRPSSFQNN 965
Db 967 RYRCYMWLPHLGGIFIEGLPESAEPTEEAALGNDIEASSVKTEKDSFGSLCKTSGHPN 1026
QY 966 VQSQDPQVST-----KTGEP-----LMSESTSNID 990
Db 1027 STAPEQNSTSCHCSDSKDEKPSGVTQFPNSVPLTNNOQLSQSVFLSWLTNKTNSIMD 1086
QY 991 -----QGPRDHSVQLP-----KPVHKPNRWCFYSCEQDLDLIEALNSRGHRE--SALKE 1038
Db 1087 STVLTPSSPHSESTPIISSEATEKPEQWL-----PLISRTCPRNHNOGLSTHSS 1137
QY 1039 TLLOEKSRICA-----QLARFSEKHFHSDKPPQPSKP---TYSRGRSSNA----- 1081
Db 1138 NRLSPSPSTAATSVKQVNEFTTEAQTATSLPSNSTPCHVCVNSGKSTASHEITLTSNI 1197
QY 1082 -YD-----PSQCAEKQLELRLRDLFD-----IED----- 1106
Db 1198 LHDSEKERRGRPPSKLL--KQIEQYNYQLIERPIAGVRQKWWMIKDPAMLESLLKAL 1255
QY 1107 -----RIYQGTGLGAIKVTDRHI-- 1123
Db 1256 HPRGIREKTLHKYLTKLHLQHLKEMCARPASDALFKFTVPDGHRSVQETLDRKSVTDLTFQ 1315
QY 1124 -----WRSALSGRYELLSEENKENGIIKTVNEDVE-----EMEIDEQTKVIVKORL 1170
Db 1316 VDSALQWVEDLEQ---RVMLSDIQRCQWSPSPADSVRTDLKYEHQLEPADDITVK--- 1369
QY 1171 LGITETPSTVSTNASTPQSVSSVVHYLAMALFOIEQGIERRFLKAPLDASDGRSYKTV 1230
Db 1370 --VKRE--DCRLYRESSNPLDLA-----VLRILCLEENVERKYLKEPLWLFSEVQHEKV 1420
QY 1231 LDR-----WRESLLSSASLSQVFLHLSLTLDRSVIWSKSLNARCK 1270
Db 1421 ITNPEDPLSTTEIEYSTSLRLRLMRQTVRCRRAQSLCLOQLERSTAWERSLKNVKTCL 1480
QY 1271 ICRKKGDAENMLVCDGCDRGHHTYCVRPKLTVPEGDFCEPCRPKQRCRLSFRQPSL 1330
Db 1481 YCRKGDNDLILLCDSCDRGCHTYCHRPWNIEPEGDFCPTC-----LSL 1526
QY 1331 ESDEVEDSMGGEDEVDGDEEBEQSEEBEVEQDEDDSQSEEBEVSUPKRGPPQVRLPV 1390

```

```

Db 1527 QS-----ESEF----- 1532
Qy 1391 KTRGKLSFSFGQQPGYPSRSQOSTPKTTVSSKTSRLKINSAPPTETKSLRIA 1450
Db 1533 -----LRSSGSK-----RIRKCTVRFTEDESPS----- 1555
Qy 1451 SRSTRHSHGQLQADVVELLSPRKRRGRKSNANTPENSFPNPRVIATKSSSQSRSVN 1510
Db 1556 -----NPSRRRHPTASQSPGESP----- 1575
Qy 1511 IASKLSLQESSKRCRKRQSPSPSVTLGRRSSGRGGVHLSAFQLVVELVRHDDS 1569
Db 1576 -ASK-----KRMGRTSQSP-----DLTFCEIILMELESHEDA 1607
Qy 1570 WPFLLKLSKIOVPDYDIKKPIALNIIREKNKCEYKLASEFDDIDELMESNCFEYNPR 1629
Db 1608 WPFLEPNPRVLPGYRKIKRPMDFSTMRHKLNGNYSRCPEAFDAELIFSNQCLFNEF 1667
Qy 1630 NTSEAKAGTRLQAFF 1644
Db 1668 ESDVGKAGLILKXFY 1682

RESULT 21
BA2A_MOUSE STANDARD; PRT; 1850 AA.
AC Q91Y55;
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Bromodomain adjacent to zinc finger domain 2A (Transcription
DE termination factor-I interacting protein 5) (TTF-I interacting protein
DE 5) (Tip5).
GN Name=Baz2a; Synonyms=Tip5;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21423468; PubMed=11532953; DOI=10.1093/emboj/20.17.4892;
RA Strohn R., Nemeth A., Jansa P., Hoffmann-Rohrer U., Santoro R.,
RA Laengst G., Grummt I.;
RT "NoRC-a novel member of mammalian ISWI-containing chromatin remodeling
RT machines";
RL EMBO J. 20:4892-4900(2001).
CC -!- FUNCTION: May play a role in transcriptional regulation
CC interacting with ISWI. May serve a specific role in maintaining or
CC altering the chromatin structure of the rDNA locus.
CC -!- SUBUNIT: Together with ISWI/SNF2h, it forms a complex termed NoRC
CC (nucleolar remodeling complex).
CC -!- SUBCELLULAR LOCATION: Nuclear. Colocalizes with the basal RNA
CC polymerase I transcription factor UBF in the nucleolus.
CC -!- SIMILARITY: Belongs to the WAL family.
CC -!- SIMILARITY: Contains 4 A-T hook DNA-binding repeats.
CC -!- SIMILARITY: Contains 1 bromodomain.
CC -!- SIMILARITY: Contains 1 DDT domain.
CC -!- SIMILARITY: Contains 1 methyl-CpG-binding (MBD) domain.
CC -!- SIMILARITY: Contains 1 PHD-type zinc finger.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AJ309544; CAC69992.1; -.
CC HSSP; Q9UIG0; 1F62.
CC MG; MG1; 2151152; Baz2a.
CC GO; GO:0005731; C:nucleolus organizer complex; ISS.
CC GO; GO:0003677; F:DNA binding; ISS.

```

```

DR GO; GO:0030528; F:transcription regulator activity; ISS.
DR GO; GO:0006338; P:chromatin remodeling; ISS.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; ISS.
DR InterPro; IPR000637; A-T_hook.
DR InterPro; IPR001487; Bromodomain.
DR InterPro; IPR004022; DDT.
DR InterPro; IPR001965; Znf_PHD.
DR Pfam; PF02178; AT_hook; 4.
DR Pfam; PF00439; Bromodomain; 1.
DR Pfam; PF02791; DDT; 1.
DR Pfam; PF01429; MBD; 1.
DR Pfam; PF00628; PHD; 1.
DR PRINTS; PR00503; BROMODOMAIN.
DR SMART; SM00384; AT_hook; 4.
DR SMART; SM00297; BROMO; 1.
DR SMART; SM00571; DDT; 1.
DR SMART; SM00391; MBD; 1.
DR SMART; SM00249; PHD; 1.
DR PROSITE; PS00633; BROMODOMAIN_1; FALSE_NEG.
DR PROSITE; PS50014; BROMODOMAIN_2; 1.
DR PROSITE; PS50827; DDT; 1.
DR PROSITE; PS50982; MBD; 1.
DR PROSITE; PS01359; ZF_PHD_1; FALSE_NEG.
DR PROSITE; PS00166; ZF_PHD_2; 1.
KW Bromodomain; Coiled coil; DNA-binding; Nuclear protein; Repeat;
KW Transcription regulation; Zinc-finger.
FT DOMAIN 499 570 MBD.
FT DNA_BIND 602 614 A-T hook 1.
FT DNA_BIND 623 635 A-T hook 2.
FT DOMAIN 800 865 DDT.
FT DNA_BIND 1137 1149 A-T hook 3.
FT DNA_BIND 1351 1363 A-T hook 4.
FT ZN_FING 1623 1673 PHD-type.
FT DOMAIN 1755 1825 Bromodomain.
FT DOMAIN 613 738 Lys-rich.
FT DOMAIN 647 774 Coiled coil (Potential).
FT DOMAIN 1006 1012 Poly-Glu.
FT SEQUENCE 1850 AA; 205597 MW; E5DD4FED1D46DAE1 CRC64;

Query Match 7.6%; Score 657.5; DB 1; Length 1850;
Best Local Similarity 20.9%; Pred. No. 6.8e-19;
Matches 314; Conservative 191; Mismatches 427; Indels 569; Gaps 54;

Qy 424 RRRGPPKRIHISQEDNVANKQTILASYSRKATKRDKLLKQEEEMKSLAFKAKLKREKAD 483
Db 624 RGRGRPPK-----IKPELLNK---TDNRLPKKLETTQILSDDDKAKMTKNGKM-RQKV 675
Qy 484 ALE-----AKKKEKEDKE--KKREELKKI-VEEERLKKKEKERLKVREKER----- 528
Db 676 RGSQTPVQGOARNKRKQDTKSLKQKDTKKLKAEEKMTKQEKLEKVKREKKEKVA 735
Qy 529 -----EKLSEERKRVYVYKQMSKPRDEMCDDLKELPFTPVK- 567
Db 736 KGEGPRAPRSCRADTKLATQKRLSEQOQQAILEEMKKPTEGMCLSHDQPLPDTFIPG 795
Qy 568 TRLPPEIFGDALVLEFLNAPGEL--FDLQDEFDPDGTVLEVEALV-GNDSEGPLCELL 624
Db 796 LTLSSRAFSDCGLTIVFLHSFGVLGDFLTKDVP---SLGLVQLQGLLCQGSLSLKVQDLL 852
Qy 625 FFFLTALFOAIAEBEEVAXEQLTADTKGSKSLDLDCTLSLILRLHLTLAGADVTS 684
Db 853 VRLKAL-----HDPGLPPYCQSLKILGKEMSEIPLTRDNVSEILRCFLMA----- 899
Qy 685 ANAKYRYKRGGFGFATDDACMELSNPSLVKLSSTSVYDLTFPGERKMILHALCGKLLT 744
Db 900 ---YRVE-----PPFCDSLRTQPFQAPQPPQQAAILAFLVHEL-- 934
Qy 745 LVSTRDIEDYVDILROAKQEFRELK-AEQHRKEREREAAIRKKEELKEQEKM-K 801
Db 935 --NSSTIIINEIDKLTLESVSCRKNKRVIVEGRRLRLKTALAKRTGRPEVMMEGADGLGR 992
Qy 802 EKQEKLEDEQRNSTADISIGEEREDFDTSIESKDETEQKELDQDMFTEDDDPGSHKRG 861

```

Db 993 RRRSIMEETS-----GIEEEEEENTTAV-----HG 1019
QY 862 RRGKRGQNGKEFTROEQINCVTRRELLTADDEE-EALKQBHQKKEKELLEKIQSALACTNI 920
Db 1020 RRRG-----KEGEIDVAASSTPELERHIEKLSKQLFRKLLHSSQMLRAVS-- 1067
QY 921 PFLGRDRMYRRWIPSPISGLFIE-EDYSGLTEDML----- 955
Db 1068 --LGQDRYRRHWLVPYLAGIFVEGSEGSTVTDEIKQETESIMEVVTSTPSSARASVKR 1125
QY 956 -----LRRPSFQ-----NNVQSQ-----DPQVSTKRG 978
Db 1126 ELTGNSASTSPARSGRPKPKGSLQPOHLOSTRECEAOQTVHPPOQLOAPF- 1184
QY 979 EPLMSESTNIDQGRDHSVQLPKPVHKPNRWCFFSSCCQLDQLIEALNSR-----G 1030
Db 1185 QPHLQPSGFLF--PEGSPFSLGQSOHDLQSQAFLSWLSQTOHNSLSLSSVLTPDSSPG 1242
QY 1031 HRESALKETLLQEKSRICQALARFSEKPHFS-----DKPOP----- 1067
Db 1243 KLDAPSQSLEBPEDEAQCPCGPGQWENFSAQIPCDAAPTPPPAVSBDQPTPSLQLLA 1302
QY 1068 DSKPTYS-----ELPALEEG-----VMSWSPE-----KTYETDLAV 1464
Db 1303 SSKPMNTPGANPCSPVQLSSTHLPFGTPKRLSGDSEMSQSPTGLGOPKRRGRPPSKFF 1362
QY 1083 -----DPSQMC-----EKOLELRL-----RDF 1100
Db 1363 KOVEQHLYLTQAQIPPEMCGWWMWRDPETDLVLLKALHPRGIREKALHKLHSHKHDF 1422
QY 1101 LLDI-----EDRIYQGLGAIKVTDRHIWRSALSGRYELLSEENKENGIIKTVNED--- 1152
Db 1423 LQEVCLQPLTDRIPEPN-----ELPALEEG-----VMSWSPE-----KTYETDLAV 1464
QY 1153 ---VSEMEIDEQTKVIVKD-RLLGKIKTETPSTVSNAS-----TPQSV----- 1191
Db 1465 LQWVELE-----ORVVLSDLOIRGWTCTPTDSTREDLTYCEHLPDSPEDIPWRGREGT 1520
QY 1192 -----SSVHVYLAVALFOIEGIERRFLKAPL-----DASDSGR 1225
Db 1521 VPQRQNNPLDLAVMLAVLEQNVERRYRLREFLWAHEVVVEKALLSTPNGAPDGTSTET 1580
QY 1226 SYKTV--LDRWRESLSSASLSQVFLHSLTDRSVIWSKILNARCKICRKGDAENMVL 1283
Db 1581 SYEITPRVVRVWQTLERCRSAQVCLCMGQLERSIAWKSVMKVTCVLCRKGDNDEFLL 1640
QY 1284 CDGCDRGHTYCVRPKLTVPEDWPCPCBPCKQRRLSFRQRPSSLESDEVEDSMGGE 1343
Db 1641 CDGCDRGCHYCHRPKMEAVPEGWFCVCL----- 1671
QY 1344 DDEVDGDEEGOSEEEVEDEDEDSQEBEEVSLPKRGRPOVRLPVKTRGKLSSESFSR 1403
Db 1672 -----SQVEEY-----TQRPFPKRGQ-----KRKSPF--- 1696
QY 1404 GQOQBFGRYPSRQOSTPKTTSVSKTGRSLRINSAPPTETKSLRIASRSTRSHSGPLQA 1463
Db 1697 -----PLTFPE-----GDSRR-----RMLSRS----- 1713
QY 1464 DVFVELLSPRKRGKGSANTPENSPPNPRVIAATKSSQOSRVNIASKLSQESBSK 1523
Db 1714 -----RDSAPVRY----- 1722
QY 1524 RRCRKRQSPSPVTLGRSSRGQGVHLSAFEQLVVELVPHDDSWBPLKLVSKIQVPD 1583
Db 1723 -----PEDGLSPKRRHRMRSHHSDLTFCCEILMEMESHDAWPFLEPNPRLVSG 1774
QY 1584 YYDIIKPTIALNIIKRVNKCYSKLASEFIDDIELMFNSCNFPNPNRNTSEAKAGRLQAF 1643
Db 1775 YRVVKNPMDFTMERLLRGYTSSEFEFAADALLVFDNCQTFNEDDSEVGKAGHVMRRF 1834
QY 1644 F 1644

Db 1835 F 1835
RESULT 22
QY 0703S9 PRELIMINARY, PRT; 2060 AA.
ID Q703S9
AC Q703S9;
DT 01-MAR-2004 (TREMBLrel. 26, Created)
DT 01-MAR-2004 (TREMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
DE AGCP10911.
GN Name=agCGS0231; ORFNames=ENSANGG000000009298;
OS Anopheles gambiae str. PEST.
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Anopheles.
OX NCBI_TaxID=180454;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PEST;
RA Anopheles Genome Sequencing Consortium;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AAAH01008964; BAA12387.1; -.
DR HSSP; Q9UIG0; 1F62.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR001487; Bromodomain.
DR InterPro; IPR004022; DDT.
DR InterPro; IPR001739; Methyl-Cpg_bind.
DR InterPro; IPR001965; ZnF PHD.
DR Pfam; PF00439; Bromodomain; 1.
DR Pfam; PF02791; DDT; 1.
DR Pfam; PF01429; MED; 1.
DR Pfam; PF0628; PHD; 2.
DR PRINTS; PR00503; BROMODOMAIN.
DR PROSITE; PS50014; BROMODOMAIN_2; 1.
DR PROSITE; PS50827; DDT; 1.
DR PROSITE; PS01359; ZF PHD 1; UNKNOWN_2.
DR PROSITE; PS50016; ZF PHD_2; 2.
DR SEQUENCE 2060 AA; 228520 MW; CBFF2DDFF7C95D40 CRC64;
Query Match 7.1%; Score 614.5; DB 2; Length 2060;
Best Local Similarity 18.6%; Pred. No. 4.5e-17;
Matches 362; Conservative 267; Mismatches 542; Indels 777; Gaps 70;
QY 305 SVDETIISDSDSETQSCSFQNGKKDAIDPLLFKYKVQPTKXELHESAIKATQISR 364
Db 270 SMDNST-----SNTNTDTSISGSGMSEGEHINLRELVRPVEKGRWRRTVIRG--LTR 323
QY 365 RKHLFSRDKLFLKQHCPEQGV--IKIKASSLSTYKIAEQDFSYFFDDPTTFISPA 422
Db 324 NGHI--KGDVYYPQSVNKMGMNQIQLYLDQFKPKDLSRDNFSFSAKAIIVGFL-QPA 380
QY 423 NRRGRPRKRIHISOEDNVANKQTLASVRSKA---TKERDKLLKQEMKSLAFKAKLKR 479
Db 381 PLPVATDEGFKMTDVEVARRELDLKMTRHAGLVGEQRIEIAKQQQALR---DAKKLAK 437
QY 480 EKADALEAKKKEKEDKEKKRELKKIVBEERLKKKKEKERLK-----VEREKEREKLEBK 535
Db 438 EEMN---KNKEKE--RERRRQHMALIKQELRRRKFEKKEKKHQVLDKLIQREKKLVNR 492
QY 536 RKYVEYLLKQNSKPRMEDCDLDELPEPTPVK--TFLPPEIFGDALMLVLEFLNAFGL--F 592
Db 493 KRDNTNLAELRKPOEDSEIVDQTVLPFSRIPGLKLTGTGYADLLMVFLEFLNFGTGLG 552
QY 593 DLQDEFPDGVTLVLEALVGNDS-----EGP----- 619
Db 553 DM-ESLPN---LQSHLALTSENAIDAEELLSSVMTHLLVCAIEDPGIPNPGRHTLLGQ 608
QY 620 -----LCELLFFFLTAIFQA-----IAEEEEVEAKEQL-TDADTKGCSLK 658

Db 609 TLRQADITHNVSEILRIYLYAVATGEVKLQSGINLERDRDAPSKHLLVNDDEDFQMCSTK 668
Qy 659 S-----LDLDSCTLSLSEILRL-----HILASGADV 682
Db 669 NSQFYELLAENARYKLSBLKDKPFVALNPTTKTEILAMLCNDLLMNKAVCKQIDSSLEA 728
Qy 683 TSANAKYRY-----QKRGGF-----697
Db 729 QAOLKKERYLDNKRKYKMLSPVKMGVEVGGETPNGGVFNIPGKPLTAEMWTNGEQPSQ 788
Qy 698 -----DATDDACHMELRLSNPLVK-----KLSSTSVYDL 726
Db 789 HQQTFQVHQNSNNSASELDDLKSMINDHSTCGDRADEDSLSESGTQLEDEDAHL 848
Qy 727 TPGE-----KMKILHAL-----CGKL-----742
Db 849 TAEAAQRKYDKILETSFONKQOLENALNQLRVKCFGQDRYWRYYNGLGCKGCIYVEAMES 908
Qy 743 -----LTLVSTR-DFIEDYVDILR--QAKQEFRELKABOHRKEREAAA---783
Db 909 TQPEMYRIENALEEVOSRPDVPKPAKIPESAVKQDPAEDVAABCKKEGEVDVAVKPE 968
Qy 784 -----RIRKRKE-----EKLKEQEQMKKEQE-----KLKE 809
Db 969 IIAADAQGTQBELQPTEDRKRKRCSSVSLKLLKKKKKQRTTSEGGESEFRNLGSAATPE 1028
Qy 810 DEQ-----RNSADISIGEEREDFD-----TSIESKDTQK---841
Db 1029 DDDVMTVQEBETPTITIPDDDTESGDRMPTAPCEGVKTEHDNKGSTTLTDSMNCMDKPKL 1088
Qy 842 -----ELQDMFTEDDEDPG--SHKGRGK-----RGONG-----870
Db 1089 MENGVMRDPGELVEQCQOEVDDBEIVTGARGNHGSANDCKPMIAAGDNGVDRKPRTT 1148
Qy 871 -----FKETROEQINCVTRELLTADDEEALKQEHORKE-----906
Db 1149 TTTSSVISGSDVEMIEVKDESNCAVRAPITPGSYLYLRNTDKQEFKESDQLLNRFW 1208
Qy 907 -----LLEKTSQAIACNIIPLGRDRMYRYW-----I 934
Db 1209 SIVDKELPLSCEPLPTINGSTPASLARQIFTNITCREICQIQGNR-----WDIGNNIQ 1263
Qy 935 FPSIP--GLFIE-----EDYSGLTEDML-----LPRSSFO-----963
Db 1264 FFSVLEKGVLEHFNESILMSGLDDDEINEVIAKSRPPLQSDMKPAFKRETIYS 1323
Qy 964 -----NNVQSDQPVSTKTGEP-----LMSESTSNID-----QGPRDH 996
Db 1324 YSQHHPNQRIRAEWMAEETADPEYGNFSLPAYMTLTLNLTAIVQCQDQFPLQMTPEE 1383
Qy 997 SVQL-----PKPVHKPNR--WCFYSSCEQLDQLEALNSRGHRESALKETLL 1041
Db 1384 EKQLEDVQKHGAPOKTEPQVVPREPRYGMWKINDIEELNELIKALNPRGVRERLLRQSL 1443
Qy 1042 QEKSRIQAARFSEKEKHFSDKQPDQSKPT-YSRGRSSNAVDPDSQMAEKOELRLRDF 1100
Db 1444 ES-----LAESVNLTTP--HHVSHPRAAPPPNGYIPEAWNANPNS--IARVEVALLDQ 1494
Qy 1101 LLDIEDRIYQGTGLGAKVYTRDHIRMSALESQRYELLSEENKENGIIKTWNBDVBEIMEIDE 1160
Db 1495 IEAMEDKVASASM-QVKG-----WQMPQREG-----DSENGVV-----EDV-----1529
Qy 1161 QTKVJVKDRLLIGIKTETPSTVSTNASTPQSVSSVVHYLAMALFQIEQGIERRFLKAPL--1218
Db 1530 -TIEMLRERILG-----LEAAIERRYLKPLPLGI 1556
Qy 1219 -----DASDSGRSYKTVLDRWRRESLLSSASISQVEL 1249
Db 1557 NTTEAQMAMIAQESHSHSNQNVNLSNCSNDAEDENLPKGLLSWRDAVERSVTTAQLUM 1616
Qy 1250 HLSTLDRSVIWSKSTLNARCKICRKGDAENNVLCDCGDRGHHTYCVRPKLKTVPEGDWF 1309
Db 1617 ALYVLESCVANDKSIIMKANCQFCQSEGEDKLLLCDCGDRGVHTYCFKPRMDKIPDGDWY 1676

Qy 1310 CPECR-----1314
Db 1677 CFECKNATGDRKCIIVCGGLRPPPLGKMWYCELCPRAYHQDCYIPPMKLYPRGKWCQC 1736
Qy 1315 -----PKQRCRR-----LSFRQRPFSLESD-EDVEDS-----1339
Db 1737 VAKAPPKKKPQKPKERTNNSSQSLNSSLNQQSLNSSHEDIIATTPLSATSAAATVE 1796
Qy 1340 MGGEDDEVGDDEEGEQSEEEY-----EVEQDEDDSQEE-----EVSILPKRGORP 1384
Db 1797 QEAQSDYVASGEAAQYASAGEYTHAQSTSECEPDQESPAEVDQIEASSSPVPTAGG 1856
Qy 1385 QVRLPV-----KTRKGLSSFSRSGQQQEPGRYPSPRSQQSPKTTVTSKTSRSLRK 1436
Db 1857 EYFSPSSNANMAAGSSTSYLSDRTPGGQDGEEG-----SCGDSSEY 1900
Qy 1437 NSAPPTETKSLRIASRSTRHSHGLOADVFEVLLSPRRKRRGRKSANNTPENSFPNFR 1496
Db 1901 QPRSPVKDSLCLAGSSSSSS-----ISDHQRK-----1930
Qy 1497 VIATKSEQSVNIAKLSLOESKRRCKRQSPSPVTLGRRSSGROGGVHELASF 1556
Db 1931 -----ERSKERDEAKERAQEKATKRLK-----ELAVC 1960
Qy 1557 EQLVVELVRDSDSWPFLKLYSKIOVPDYDIIKKPIALNIIREKVNKCEYKLASEFIDDI 1616
Db 1961 KTLIEEMELHEDSWPFLPVNTKQFPYIRKVIKSPMDLSTIKKRLQDLVYKSRDFIADV 2020
Qy 1617 ELMFSNCFEYNPRNTSEAKGTRLQAPF 1644
Db 2021 RQIFDNCVEFNEDDSPVGIAGHGMRKFF 2048

RESULT 23

Q9BMQ0
ID Q9BMQ0 PRELIMINARY; PRT; 3109 AA.
AC Q9BMQ0;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)
DE Toutatis.
GN Name=tou;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21589546; PubMed=11732091;
RA Fauvarque M.O., Laurenti P., Boivin A., Bloyer S., Griffin-Shea R.,
RA Bourbon H.M., Dura J.M.;
RT "Dominant modifiers of the polyhomeotic extra-sex-combs phenotype
RT induced by marked p element insertional mutagenesis in Drosophila.";
RL Genet. Res. 78:137-148(2001).
DR EMBL; AF314193; AAK00302.1; -.
DR HSSP; Q9UIG0; 1F62.
DR Flybase; FBgn0033636; tou.
DR GO; GO:0003677; P:DNA binding; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR000194; ATPase_a/bcentre.
DR InterPro; IPR001487; Bromodomain.
DR InterPro; IPR004022; DDT.
DR InterPro; IPR001739; Methyl-CpG_bind.
DR InterPro; IPR001965; Znf_PHD_
DR InterPro; IPR001876; Znf_RanGDP.
DR Pfam; PF00439; Bromodomain; 1.
DR Pfam; PF02791; DDT; 1.
DR Pfam; PF01429; MBD; 1.
DR Pfam; PF00628; PHD; 2.
DR PRINTS; PR00503; BROMODOMAIN.


```
Qy 1268 ----- 1267
Db 2504 INKPENQAPLSIKINLKALAQNGCLLTKTPPIILTKSPASPSHPNTSDSDPGKRT 2563
Qy 1268 -----RKICRKGDAENNVLCDCGRGHHTYCVRPKLKT 1302
Db 2564 KKGSGKRRRSANNTSSKYNLQNCQCTSGENEDKLLLCDCGCKGYHTYCFKPKMDN 2623
Qy 1303 VPEGDFWPEC-----RP 1315
Db 2624 IPDGWICYECUNKATNERKIVCGGHRPSVPGKMIYCDLCPRAVHADCYIPPLIKVPRG 2693
Qy 1316 KQRCRLRFRQP-----SLESDEVEDSMGDEDEVDGDEEE----- 1353
Db 2694 KWYCHGCTSRAPPKRSAGGTSKSSKRRDRDRESGSAKRRSDNSKTPAMEHMQQQ 2743
Qy 1354 -----GQSEBEYEVEQDEDSQBEEVSLPKRGRPOVLPVKTRGKLSSFSRGGQQOE 1408
Db 2744 MPLAGGDGHHHHQPPSLNSHDSMNSLP-AGSSEV--SVGACGRLNGRRSTRSGFQ- 2799
Qy 1409 PGRVPSRSQQTPKTVVSKT-----GRSLRKINSAPPTETKSLRI----- 1449
Db 2800 ---LNNASALCSPAHSVVSATNYDDQHHANNSVDGSSRFHAHLTPPSNNGTAALLLEDVPG 2856
Qy 1450 ---ASRSTRHSHGFLQADV-----VELLSRP-----RKRGRKRSAN 1483
Db 2857 GANVMPGVYVYTPVAGNFAGLINQAPVQAPMPFANVAVMSPAVTPTRTPPTPPA 2916
Qy 1484 NTP-----ENSPNFPNFRVIATKSSQSRSVNIASKLS----- 1516
Db 2917 PTPPPPTPLMQASPTATALHNACQSPQQAQLMTWPSFAIGVGTATNQMSPPP 2976
Qy 1517 -----LQSESKRRCRKEQSPSPVTLGRSSRGQGVHLSAFEQLVVELVRHDSWP 1571
Db 2977 INIHAIQAKELKQEKKEKATKKL-----MKELAVCKTLGEMELHEDSWP 3024
Qy 1572 FLKLVSQIVPDYDIIKKPIALNIIREKNVCKEYKLAISEIDIELMFSNCFYVNPRT 1631
Db 3025 FLPLPNTQFTYRKIIKTPMDLSITIKKLQDLSYKTRDEFCVDVQIFDNCMFNEDDS 3084
Qy 1632 SEAKAGTRLOAFF 1644
Db 3085 PVKGAGHGMKRF 3097
RESULT 24
Q9V602 PRELIMINARY; PRT; 3080 AA.
AC Q9V602;
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-OCT-2002 (TRENBLrel. 22, Last sequence update)
DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)
DE CG10897-PA.
GN Name=tou; ORFNames=CG10897;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
RA Abril J.F., Agbayani A., An H.J., Andrews-Franckoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
```

```
RA Burtis K.C., Buesam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delvey A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrera S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glaeser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.B., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klamis I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.Y., Wassarman D.A., Weinstock G.M., Weissenbach J., Ye J.,
RA Williams S.M., Woodager, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster";
RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426065; PubMed=12537568;
RA Celniker S.E., Wheeler D.A., Kronmiller B., Carlson J.W., Halpern A.,
RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,
RA George R.A., Hoskins R.A., Laverty T., Muzny D.M., Nelson C.R.,
RA Pacleb J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,
RA Svirkas R., Taber P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
RA Weinstock G., Scher S.E., Myers E.W., Gibbs R.A., Rubin G.M.;
RT "Finishing a whole-genome shotgun: Release 3 of the Drosophila
RT melanogaster euchromatic genome sequence";
RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426070; PubMed=12537573;
RA Kaminker J.S., Bergman C.M., Kronmiller B., Carlson J., Svirkas R.,
RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,
RA Ashburner M., Celniker S.E.;
RT "The transposable elements of the Drosophila melanogaster euchromatin:
RT a genomics perspective.";
RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426069; PubMed=12537572;
RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
RA Betencourt B.R., Celniker S.E., de Grey A.D., Drysdale R.A.,
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA Lewis S.E.;
RT "Annotation of the Drosophila melanogaster euchromatic genome: a
RT systematic review.";
RL Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).
RN [5]
RP SEQUENCE FROM N.A.
RX FlyBase;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE FROM N.A.
RX FlyBase;
RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE003825; AAF58638.2; -.
DR HSSP; Q9UIG0; 1F62.
DR FlyBase; FBgn0033636; tou.
```


DR GO:0003677; F:DNA binding; IEA.
DR GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR000194; ATPase_a/bcentre.
DR InterPro; IPR001487; Bromodomain.
DR InterPro; IPR004022; DDT.
DR InterPro; IPR001739; Methyl-CpG_bind.
DR InterPro; IPR006162; Pplantn_S.
DR InterPro; IPR001965; Znf_PHD.
DR InterPro; IPR001876; Znf_RANGDP.
DR Pfam; PF00439; Bromodomain; 1.
DR Pfam; PF02791; DDT; 1.
DR Pfam; PF01429; MBD; 1.
DR Pfam; PF00628; PHD; 2.
DR PRINTS; PR00503; BROMODOMAIN.
DR SMART; SM00297; BROMO; 1.
DR SMART; SM00249; PHD; 2.
DR SMART; SM00547; Znf_RBZ; 1.
DR PROSITE; PS00152; ATPASE_ALPHA_BETA; UNKNOWN_1.
DR PROSITE; PS00827; DDT; 1.
DR PROSITE; PS00012; PHOSPHOPANTHETHEINE; UNKNOWN_1.
DR PROSITE; PS01359; ZF_PHD_1; UNKNOWN_1.
DR PROSITE; PS00016; ZF_PHD_2; 2.
DR PROSITE; PS00016; ZF_PHD_2; 2.
SQ SEQUENCE 3080 AA; 330285 MW; 32A4D6066F071AB3 CRC64;

Query Match
Best Local Similarity 17.6%; Score 591; DB 2; Length 3080;
Matches 455; Conservative 275; Mismatches 688; Indels 1160; Gaps 86;

QY 3 DASESRGVAFLNNVL----PGSPLSLPVSTGCKSHRVKANKVKEARSEKLLPTALPPS 59
DB 715 DNATNSGGV-----NLSLKSNAGSALTSPGAVGASNPITIDDPADP-----LNL5MKPS 765

QY 60 EPKVQKLPSSRRGGGGTQFPARSRAVAG-EAAARG-----AAGPER 104
DB 766 D-----KNSSSSNAAGSSSSSALANLASDYQAASGGSGNSLSITAAALGGTG 820

QY 105 SPLGRRVSPRLCSGEGGQVAVGIAGKRRGRDGSRRAPGGREPLLHRKPFVRQPK 164
DB 821 GMPGGS-----ISGSGTSPAPAGAGGATGGGSGGSGGSS-----YKEGR 865

QY 165 PADL-----RPDEEVYCKVTNEIFRHYDDFFERTILNSLVWSCAVTGRPGLTVOEAL 219
DB 866 PNLGRGVSKP-----KNTVASLLAQSRAGVLKPLMATQQLQ 904

QY 220 SEKKARQLQSPPELIPVLVLTSLTHRSRLHEICDDIFAYVKDORYFVEETVIRNNG 279
DB 905 -----QG 906

QY 280 ARLOCTILEVLPPSHQNGFANGHV-NSVDGETIII-SDSDSETQSCSFQNGKKDAIDP 337
DB 907 ADIEKIRLAL-----SEANAHMETSTDSVSAESGLSESESDAILNVAELRV--P 957

QY 338 LFKYKVOPTKELHESAIKATQISRRKLFPSRKLFKLFKQHCPOE-----GVTKI 391
DB 958 LELGW-----KRETVIRGLTKOGQI-----RGEVYYAPGSTTPLKSNQGVPAILEQ 1004

QY 392 KASSLSSTYKIAQDFSY-----FFPDDPPTETIFSPANRRGRPPKRIHISQED---- 439
DB 1005 QPSNLS-----RENFSARAIVGSLQFAPPPY-----ANDGETIRMTDSDVAKR 1050

QY 440 ----NVANKQTL-----ASYRSKATKERDKLLKQEEWKMSIAFEKAKLREKADALEAK 488
DB 1051 LEDLKVFTQTLNVESQRIEIAKQQAMRDACKLQKEE---LARNKEKARQEKSKLEQQ 1106

QY 489 KKEK-----EDKEKKREELKKTIVBEERLKKKEKGRKLVKEREKEREKURE-----EKR 536
DB 1107 RKDKELKNQQAVERERRRRQHMSLRMLELRKKFEDREKKHQVLVLDRLLLRRERRMAERK 1166

QY 537 KYVEYLKQWSKPRDEMCDDLKELPEPTPVK--TRLPPEIFGDALMVLFLNAFGL--FD 593
DB 1167 RDAEILQIRRNEDSEMPQELVPIPELDRIAGNRLPGQAMADLLMVFFLHNFGETLGF 1226

QY 594 LQDFPFDGVTLEVLVEEALVGNDSSEGLCELLFFFTTAIFQAIABEEVEAKEQLTDADTK 653
DB 1227 M-ESLP---SLQNLHDALM-SDSNADABEELLVSMTHLL--VCAIEDPGVPNPGRHTTILL 1279

QY 654 GCSLSKSLDSDCTLSEILRLHLILASGAD-----VTSANAKYVRXKGGFDATDDACM 705
DB 1280 GQSLRNADITNSNVSEILRIYIYATATGEVRQMHCITVDRRERRVDPDHQOLD-SDTTTH 1338

QY 706 ELRLSNPLSVKCLS-----STSVYD-----LTPGEKMKILHALCGKLLTLIVSTRDRFIE 753
DB 1339 SHSVKNQYVYKLLHENDTWKLSQSILKDRPFVALLNPTKRAQMLAHLNCLLNNKAVLRQID 1398

QY 754 DYVDILROAKOE-----FRELKAEQHRKEREEAAAIRKRKEKLEKEQEQMKKEK 806
DB 1399 GSLETCAQMRKEKYMMDKVRKYKALHMRKARIEAVERAQAEREAAQMAQKQKLDABR 1458

QY 807 LKEDQERNSTADISI-----GE-----EREDPDTSIE--SKDTQK-----EL 843
DB 1459 LKABEEAKAAAAEAAAAAGTGDGATKGGSPNGEKPEDGQNEGAAKEPQOQOQOPMEV 1518

QY 844 D-----QDMFTEDDDPGSHKRRGRKRG----- 867
DB 1519 DGVVDEASLVSPAKTIIQTDNSLTPSKQDMPTPTTYQINGSSSTPTTSGVTGDMNVLLQAK 1578

QY 868 QNGFKFTRQE---QINCVTREL-----LTADDEEALKQEHORKE----- 904
DB 1579 KSGARNSINDEHHHDVSIIDDDLSLDSEITNVEDEDNLSADELQKLDKIVRASLNC 1638

QY 905 KELLEKI--QSAIACTNIFPLGRDRMYRYWIFPSIPGLFTEE-----DYSGLTFE- 952
DB 1639 KEALEKSTNQURAAC-----FGQDRFWRRYWKLPKAGGIFIEALESANDICDYHEAL 1693

QY 953 -----DMLLPRPSSPQNNVQ 967
DB 1694 MDDKQDANDEKENSENEKDVAAESSEQPMVEYDESTITKLEDGVPASDVCMPE-SNQQNAHQ 1752

QY 968 SQD-----POVSTKTGEPMLSE----- 984
DB 1753 DEEDDDDDVTINKVPEIVDLGDDDDAAPPKIEPPRPEIKVKSEMELMGPPTTMI 1812

QY 985 ----- 984
DB 1813 STKTDFAEAIKIPMPGILMPPTLNNTNNNNNGSDNCDKLETGLGLGQOQQNFQS 1872

QY 985 -----OLP----- 1001
DB 1873 VIKTEDVKKDDCIIIVSTSSVDDTPKWFISIVRREVPLISELPABEGGVVGOELOISYANQ 1932

QY 1002 -----KPVHKPNRWCYS-----SCEQLDO----- 1021
DB 1933 NCSAQQLQGHPPDLINNMQYYSIPMDCKVDTSKLGNECIFSLSLGDEKQMLAKVEYK 1992

QY 1022 --LIEALNSRG--HRESALKETILQEKSRICAQLARFSE-----EKPH--PSDKPQ 1066
DB 1993 AHKVESKNGLSGPHRHHTKDDDEQAKLKDKETDTEMETDADDLAGKEKFLRSDVP- 2051

QY 1067 PDSKPTYSRGRSSNAYDPSQCAEKLRLRDLF-----LDI- 1104
DB 2052 PDTGGVSEGTD-----VKPKIELRLDEALSQAYYHNANMSSLSVVQTYPIIDIP 2101

QY 1105 -----EDRIYQGTIGAIVT-----DRHIWRSALSGRYELISEENKE 1142
DB 2102 LPLSMTDPDEHL---LEQVKLAGFPFVRVGHVPPRRQRYGWWQ-----LDEQKL 2148

QY 1143 NGIITKTVN-EDVEWEDEQTKVIVKORLLGIKT----- 1175
DB 2149 RQLLKTLNPSGLRERELQENLQ-----RFLGLEQPLGVNYKLKSIDIDPFEEFLMPDKXGD 2203

QY 1176 -----ETPSTVSTNAS-----TPQSVSSV 1194
DB 2204 WNPVAKRVEALIEQLESLEDKVASQMLQNLQWPNRVESELTDSDQEDVTEDFVSI 2263

QY 1195 VHYLAMALFQIEQIGTERRFLKAPLDAS----- 1221

```
Db 2264 IPMIRIIDL EANIERRYLPPLGSGQDAHLAVIAQNQHTTTTQTQNSASAAVLLQM 2323
QY 1222 -----
Db 2324 QQQQQQQQAQQQQQQQGGAGNSLNPSSFNRTMALAAAAAAGPGNATGVANSVVAG 2383
QY 1222 -----DSGRSYKTV-----LDRWRESLSSASLSQVFLHLSTL 1254
Db 2394 ATPESGSGPENSGNASPASNCSDRDEKVGQIPKGLVQWRDAVRSRHTTAQLAWALYVL 2443
QY 1255 DRSVIWKSILNA-----
Db 2444 ESCVANDSIIKAVATKNKSSKKSSAKKQATPSKKQQKKKKKEQQLTPNGKSKSKSA 2503
QY 1268 -----
Db 2504 INKPENQAPLSIKINLKALAQNGCLTKTPPILTKSPSPSHPNWTSDDSDFGKRT 2563
QY 1268 -----RCKICRKGDAENNVLCDCGRGHHTYCVRPKLKT 1302
Db 2564 KKKGGKRRRANNTNSSKYNLSNLCQFCSTGENEDKLLLCDCGCKGYHTYCFKPKMDN 2623
QY 1303 VPEGDFWFCPEC-----RP 1315
Db 2624 IPDGWICYECVNKATNERKCI VCGHRPSPVGKMIYCDLCPRAVHADCIYIPLLKVPRG 2683
QY 1316 KORCRLLSFRQP-----SLESDEDEDSMGEDDEVDGDEE----- 1353
Db 2684 KWCYCHGCI SRAPPKPKRSAGCTSGSSSKSRDRDRDREGSGSARRSRSDNSKTPAMEHMQQQ 2743
QY 1354 -----GQSEEEYEVEQDEDSQEEVEVSLPKRGRPQVRLPKT-----RGKLSSESFS 1401
Db 2744 MPLAGDGHHHHQPPSLNSHSDSNLSLPAAPLSPASHVSATNYDDQHANNVSDGS 2803
QY 1402 SR-----GQQQEPGRYPSRSQQSKTITVSSKTCGRSLRKNSA 1439
Db 2804 SRFAHLIPPSNGTAALLEDVPGANVMGVYP-----VYTPVAGNFSAG-----LINQA 2855
QY 1440 PPTETKSLRIASRSTRSHGHPLOADVVELLSPPRKRGRKSNNTP-----EN 1488
Db 2856 PVQPAMPANVAVMSPR-----AVTPTRTRTPTPAPTPTPPPPPTLLMQA 2902
QY 1489 SPNPFNFVIATKSEQSRSVNIASKLS-----LQESSEKERC 1526
Db 2903 SPTALHVNACQSPPOQQAQLMTMPSPPAIGVGTATNQMSPPPINHIAQEAKEKLQ 2962
QY 1527 RKROSPEPSPVTLGRSSRGQGVHLSAFQVLVVELVRHDDSWPFLKLVSKIQVPDYD 1586
Db 2963 EKKEGHATKKL-----MKELAVCKTLGEMELHEDSWPFLLPVNTKQPPTYRK 3010
QY 1587 IIKXPALNIIRKYNKCEYKLASFIDDIELMFNSCFEYPRNTSEAKGATRLQAFP 1644
Db 3011 IIKTPMDLSTIKKLQDLSYKTRDFCVDVVRQIFDNCMFNEDDSPVGKAGHGRKFF 3068

RESULT 25
QB8CAU9 PRELIMINARY; PRT; 657 AA.
AC QB8CAU9;
DT 01-MAR-2003 (TremBLrel. 23, Created)
DT 01-MAR-2003 (TremBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TremBLrel. 26, Last annotation update)
DE Mus musculus 16 days neonate thymus cDNA, RIKEN full-length enriched
DE library, clone:A130044E01 product:bromodomain adjacent to zinc finger
DE domain, 1B, full insert sequence. (Fragment).
GN Name=Baz1b;
OS Mus musculus (Mouse);
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
```

```
RC STRAIN=C57BL/6J; TISSUE=Thymus;
RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning.";
RL Meth. Enzymol. 303:19-44(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Thymus;
RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RA RIKEN FANTOM Consortium;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Thymus;
RA The FANTOM Consortium;
RT "Analysis of the mouse transcriptome based on functional annotation of
  60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Thymus;
RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
  Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT "Normalization and subtraction of cap-trapper-selected cDNAs to
  prepare full-length cDNA libraries for rapid discovery of new genes.";
RL Genome Res. 10:1617-1630(2000).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Thymus;
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA Shibata K., Itoh M., Aizawa K., Nagao K., Nagasaki T., Tashiro H., Itoh M.,
  Konno H., Akiyama J., Nishi K., Kitsuai T., Nishino T., Harada A.,
  Yanamoto R., Nakamura S., Hazama M., Nishino T., Kashiwagi K.,
  Yamamoto R., Matsuura H., Sakaguchi S., Ikegami T., Kashiwagi K.,
  Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
  Yoneda Y., Ishikawa T., Orawa K., Tanaka T., Matsuura S., Kawai J.,
  Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
  sequencing pipeline with 384 multicapillary sequencer.";
RL Genome Res. 10:1757-1771(2000).
RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Thymus;
RA Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
  Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
  Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Kasukawa T.,
  Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kouda M., Koya S.,
  Katoh H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,
  Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
  Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,
  Saito K., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
  Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,
  Tagawa A., Takahashi F., Takaku-Akaira S., Takeda Y., Tanaka T.,
  Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
EMBL; AK037737; BAC29862.1; -.
HSSP; Q9UIG0; 1F62.
MGD; MGI:1353499; Baz1b.
GO; GO:0005721; C:centric heterochromatin; IDA.
GO; GO:0000793; C:condensed chromosome; IDA.
GO; GO:0005634; C:nucleus; IDA.
GO; GO:0005333; P:protein binding; IPI.
GO; GO:0006338; P:chromatin assembly or disassembly; IDA.
GO; GO:0006338; P:chromatin remodeling; IDA.
InterPro; IPR001487; Bromodomain.
InterPro; IPR011011; FYVE_PHD_ZnF.
InterPro; IPR001965; ZnF_PHD.
InterPro; IPR01841; ZnF_ring.
Pfam; PF00439; Bromodomain; 1.
Pfam; PF00628; PHD; 1.
```


Db 214 NNAASNMNVMWQVAAQMOQOQOQKDTQ--KKADQA-----KKAKELAKQOQKEQD 266
Qy 356 IVKATQISERKHLPSRDKLFLAKHCEPQGVIKIKASSISTYKIA----- 402
Db 267 VKNKQOEILFLMAHQHQLN-HQKHEKQKQADAAALAAKVAALAALESDESPEGKKTN 325
Qy 403 -----EQDFSYFFPDPTTFIFSPANRRRRGRPPKRIHIS 436
Db 326 EAMLRPLPLQWRQTCVRSASAGVGDVGFAPCGKKLSTYSVBYLTK--NSIHYI 383
Qy 437 QEDNVANKOTLASYRSKATKDKLLKQEMKMSLAFKAKLKR----- 480
Db 384 TRDNFLNTKLIVGEBFVPKQTEADETOQEREFAMFTEDDINKELTRLNLVFLVPKIQAS 443
Qy 481 -----KADALAKKEKEDKKEBELKKIIVEERLKK--EERLKVREKEREKLR 532
Db 444 TSNGVHEDDIKMSKIEEPDEPLDPSLNDFTTELVLHVSQIMSNVNGVDECKI-REREADLL 502
Qy 533 EEKRYVYLKQWSKPREMCDLKLPEPTPVKTR-LPPEIFGDALMWLEFLNAFCEL 591
Db 503 -----VNINDVRHLPDFSRIGNQCLSSQGFADALMVHFVQNFQHV 543
Qy 592 FDLQDEPDGVTLV-LEBALVNDSEGPLCELLFFFLTAIFQAIABEEEEVAKEQUT 648
Db 544 L-----GIDLEIAPKLESICAGLDGDANHAQOT-LQLTRQLLRALALEPFGMGNEK-- 592
Qy 649 DADTKGCSLKSLDSDCTLSEILRLHILASGADVTSANAKRYQKRGGFDDATDCAEMLR 708
Db 593 ---RFGGGGMBGLDRENFSEVMRLFLDKG-----KRGE-----E 625
Qy 709 LSNPSLVKLKSTSVYDITPGCKMKILHALCGKLLTLVSTRDFEDYVDVILROAKQEPRE 768
Db 626 LSQP-----LITCNPLSIPQKASILAFLOE---LVCSRNVTE---IDKNLDEISR 673
Qy 769 LKAEQHRKE---REBAARIRKRKEK---LKEQKQMKQEKLEKDEQORNST----- 816
Db 674 LKGEKMWREGKARALRSARSKONDEKVVVKEEQNHESDSEPTPRDTPKKATVAPPTV 733
Qy 817 ---ADISTGEEREDFDTSIESKTEQKELQDMFTEDEDDPGSHKRRGRKRGONGPKE 873
Db 734 VSVSPVSAQOQQRKF-----TPG-----LGO----- 755
Qy 874 FTRQEQINCVTRRELLTADBEBAALKE-----HORKEKLEKIQSAIACTNIFPLGR 925
Db 756 -----C---EVLTQEEKMSLQOMDSLIGDLHQ-EAQINQKIHTGTGKIRSFPPGT 803
Qy 926 DRMYRYWIFPSIPGLFIEEDYSGLTEDMLPRSSFQNNVQSOPQVSTKTGPELMSSES 985
Db 804 DRPHRYWMLAHTDKVIES-----LATTSVNPNACNANEVASKDP---PTLEQRPVGA 854
Qy 986 TSNID-----QGRDRHSVQLPKPVHKPNR-----WCFYSSCEQLDOLIEA 1025
Db 855 CETIDLDVIACVEDLVDDVLLRAKADKTRKRRRIENHMKRGWMTWQNRDCVESLRSC 914
Qy 1026 LNSRGHRESALKETLQ-----EKSRICAQLAR-----PSEBKF 1059
Db 915 MLRSQIRERALHRLTKFWFNLKFGTITTEPVGEKSDL--ELVRRQGWTRLNATIDKL 972
Qy 1060 HFSKDPQDPSKPTYSGRSSNAYDPSQMCBQKLELRDLFLDIEDRIYQGTLGAIKVT 1119
Db 973 QCHLKMSDVSKPL-----PSITPFETQKPI-----VVPPTWALAQIV 1009
Qy 1120 DRHWRSALESGRYELLSEENKENGIIKTVNEDEVEMEIDEQTKVIVKDRLLGKITTPTPS 1179
Db 1010 -----KDDMAWKVIDEVDQBELDE---TIIRQKI----- 1036
Qy 1180 TVSTNASTPQSVSSVHYLAWALFOIEGIERFLKAPLDASDSGRSYKTVLDRWRESLL 1239
Db 1037 -----IETADMDDQTCQLFEDWKSYS 1058
Qy 1240 SSA-SLSQVFLHLSTDRSVIWSKILNARCKIKRKGDAENMWLDCDGRGHITYCVRP 1298
Db 1059 TEAQTTSLMVALQTLQEGMIMWSSREALCQIC-KSMDGDEMLVCDGCSGCHMECFRP 1117

Qy 1299 KLKTVPEGDFPCPCRPKQRCRLSFRQPSLESDVEDVDSMGGEDDDDEVDEBEGQSEE 1358
Db 1118 RMTKVPEGDFPC-----QRCR----- 1133
Qy 1359 EYEVEQDEDDSQBEESVLPKRGQPVQLPVKTRGKLSSSFSSRGOQOQEPGRYPSRSQ 1418
Db 1134 -----EE-----KSGRPMCMFCSRETGNLHQ-----QRCAYHVHQEC 1168
Qy 1419 STPKTTVSSKT---GRSLRKINSAPPTETKSLRTASRSHSGHPLOADVVELLSPPRK 1475
Db 1169 DGPKEAINPETFICGHC-----QEMKQMR-----FVKRLILRSE 1202
Qy 1476 RRGRK-SANNTPENSPNPNFRV-----IATKSSEQSRV--NIASKL-----SLOES 1520
Db 1203 SEERLEDDNHAENGENTKNGHMGMGALIAIGHNQNGVKGMLKRKLEVPSTGGLPKN 1262
Qy 1521 ESKRCRKRQSPSPVTLGRRSSRGQGVHELFAFQOLVVELVRHDDSWPFLKLVSKIQ 1580
Db 1263 MNKELC-----QLMLDELVVQANALFLEBPVPKL 1292
Qy 1581 VPDYDIIKPIALNIIREKVNKCEYKLASFIDDIELMFNSCFEYNPNRTSEAKAGTRL 1640
Db 1293 VPGYMIISKPMDLKTIROKNEKLIYETPEDFAEDIELMFANCRQFNIDHSEIGRAGISL 1352
Qy 1641 QAFPHIQAOQLGLHVT 1656
Db 1353 HKFFQKRWKQDKYNFT 1368
RESULT 27
Q9FNM6 PRELIMINARY; PRT; 723 AA.
ID Q9FNM6
AC Q9FNM6;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 26, Last annotation update)
DE Arabidopsis thaliana genomic DNA, chromosome 5, P1 clone:MAH20.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
[1]
SEQUENCE FROM N.A.
MEDLINE=98069011; PubMed=9405937;
RX Kotani H., Nakamura Y., Sato S., Kaneko T., Asamizu E., Miyajima N.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. II.
RT Sequence features of the regions of 1,044,062 bp covered by thirteen
RT physically assigned P1 clones.";
RL DNA Res. 4:291-300(1997).
DR EMBL; AB006697; BAB10012.1; -.
DR InterPro; IPR004022; DDT.
DR Pfam; PF02791; DDT; 1.
DR SMART; SM00571; DDT; 1.
DR PROSITE; PS50827; DDT; 1.
SQ SEQUENCE 723 AA; 84230 MW; EE7688398916110C CRC64;
Query Match 5.6%; Score 482; DB 2; Length 723;
Best Local Similarity 22.1%; Pred. No. 4.1e-12;
Matches 212; Conservative 140; Mismatches 285; Indels 322; Gaps 34;
Qy 151 MPLLRHPFVRQKPPADLRPDEEYFYKVTWNEIFRHYDDFFERTILCNLSVWSCAVTGRP 210
Db 1 MPLKKKSHKLEPPKPNLEPQELVYQVRLTNEIFRDYQLYLKLRLNLRNVMTCKSTGKT 60
Qy 211 GLTQEALESSEKARQNLQSPPELILPVLTLSITHSRLHEICDDIFAYVKQRYFVEE 270
Db 61 SLSVEEALDSEKLAGKRVQTLPRELVAPALRIIOFSTLS-LKDLADKIATELQDCFFPGA 119
Qy 271 TVEIRNNGARLOQTILEVLPSPHQNGFANGHVNSVDGETIISDSDDSETQSCSFQNGK 330

```

120  ELYGER--DGEIHPCKILR-----IVSDGDSKQKQVEVGFLD-- 155
331  KKDAIDPLLFYKYQPTKKELHESAIVKATOISRRKHLFSRDKLKFLUKQH-CEPQEGVI 389
154  -----KNKEVDDNTVLFGE DL SWK KFPF SRNFLKFSFRDSTCHSI PMWV 197
390  KIKASSI STYKIAESQDSFYFPDPDPTTFIFSPANRRRGPPKRIHISOEDNVANKQTILAS 449
198  N-----ETDREVNGSEBEAINY-----PIEDLLPPDPDVTQRPSPSR- 291
450  YRSKATYERDKLLKQEBMKSLAPEKAKLKRKADALEAKKKEKDEKQKREBELKIVDEE 509
200  YLAKAHEITRKIPKELQNK-YVFONGELVQOQK--QEDTKGRENKRKRAENGKHVAE- 254
510  RLKKEEKERLKVREKEREKRLREKRYKVEYLVKQWSKPRDM---ECDDLKEIPEPTPV 566
255  -----ETDREVNGSEBEAINY-----PIEDLLPPDPDVTQRPSPSR- 291
567  KTRLPPBIFGDALMVLEFLNAFGELFDLODEFPDGVTLVELEALVGNDSGCLCELLFF 626
292  DFSVPMDCVGLLWVDPFCTSGFQGLHLW-----RFSLEDFENAVCHKESN---LVLIWE 343
627  FLTAIQAIAEBEEBEVAKEQLTDADTKGCSKSLDLS-----CTLSEILRLHILASG 679
344  VHASLFRFLNE-----RGDSFKALQRRSRKSKITLITWAEYL----- 381
680  AD-VTSANAKRYQKGGFDATDQACMELRLSNPSLVKKLSSTSYD-LTPGEKMKILHA 737
382  CDFLESYN-----TPDLCPFV-----GTIKR-----GHYGLLDENVKLKILRE 419
738  LCGKLLTLVSTRDFIEDYVD-----ILRQAKQEPRELKAOHR-KEREAAAARIKR 788
420  LVNHITETGMPKGBIDELVEQORHALGAARREALAEARQIREEKERSKTGEANGVLNN 479
789  KEKLEKEQOKMKEQEKLK-----EDEQRNSTADISIGEERED 828
480  RLEKKQNSAQVLESSEDSKKNESTAGGSKMGVSSNIEKSKRLTGNLYLRKHKRQM 539
829  FDTSIESKDTQEKELDOMPFTEDDDPGSHKRRGRKRGQNGKFEFTQEOINGVTRRELL 888
540  TDTKIEPKEEEDKAE-----EDEEEKG-----FSVKKO-----EVK 572
889  TADEEE-----ALKQEHQKKEKELLEKIOSAIACTNIPFLGRDRMYRRIWIFPSIGPL 941
573  SASEDEKGTERRRGPEQRQYVERE-MEKI---VIRTN--PLGKDRDYNRYWWRFSNGRI 626
942  FIEEDYSGLTEDMLLPRPSPQNNVQSDPOVSTKTGPELMSESTSNIDQGRDHSVOLP 1001
627  FVE-----NSDSEE----- 635
1002  KPVHKPNRWCYSSCEQLDQIEALNBSRGHRESALKETLLOEKSRICQALRAFSEKPH 1060
636  -----WGYYTAKBELDALMGLNRKGERELSLYTOLEIFYDRI GSTLQKTKDAH 686

```

```

RESULT 28
Q86UJ6
ID ID Q86UJ6 PRELIMINARY; PRT; 955 AA.
AC Q86UJ6;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Hypothetical protein BAZ1B (Fragment).
GN Name=BAZ1B;
GE
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCBI_TaxID=9606;
[1]
RN SEQUENCE FROM N.A.
RP MEDLINE=99063792; PubMed=9847074;
RA Wilson R.;
RT "Toward a complete human genome sequence.";
```

[illegible]


```

Db 734 -----LTFPGDSRRRLMSRSDSPAVRPYDGLSPPKRRHSMRSHSD 779
Qy 1553 LSAFEOLVVELVHDDSPFLKLVSKIQVDPYDIKKPIALNIIREKVNKCEYKLASEF 1612
Db 780 LTFCEILLMEMSHDAAWFLFVNPRLVSGYRRYIKNPMDFSTWRERLLRGYTSSEF 839
Qy 1613 IDIELMFSNCFEYPRNTSEAKAGTRLOAFF 1644
Db 840 AADALLVDFNCQTFNEDDSEVGKAGHVMRRFF 871

RESULT 30
Q62178 PRELIMINARY; PRT; 617 AA.
AC Q62178;
DT 05-JUL-2004 (TReMBLrel. 27, Created)
DT 05-JUL-2004 (TReMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TReMBLrel. 27, Last annotation update)
DE Putative DDT domain-containing protein.
GN Name=OJ1009_F08.22;
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RA Sasaki T., Matsumoto T., Yamamoto K.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF004017; BAD15468.1; -.
DR InterPro; IPR004022; DDT.
DR Pfam; PF02791; DDT; 1.
DR SMART; SM00571; DDT; 1.
DR PROSITE; PS00827; DDT; 1.
SQ SEQUENCE 617 AA; 71727 MW; 4615520664AFDECS CRC64;

```

```

Query Match 5.1%; Score 441; DB 2; Length 617;
Best Local Similarity 21.7%; Pred. No. 1.7e-10;
Matches 199; Conservative 113; Mismatches 261; Indels 344; Gaps 29;

Qy 151 MELLHRKFPVRKPADLRPDSEVYCKVTNFIHRYDDFFERTILNSLWSCAVTGRP 210
Db 1 MFLFKRKPSLEPPKDLDSKVKFQIRFTREIFRDYQDYNRLNLYRQRWSCISKGS 60

Qy 211 GLTYOALESEKKARONLQSPPELIPVLYLTSLTHSRLEICDDIPAYVKDRVFVEE 270
Db 61 NUTFEALVSEHVAHSAKQKLTETELMAPVLRMIQVSTLG-LYELVEKIYASLQEAFF-E 117

Qy 271 TVEV-IRNNGARLOCTILEVLPPSHQNGFANGVNSVDGETIISDSDSETQSCSFQNG 329
Db 118 GLELYAKQDGLAACRLKILG-----SDG 142

Qy 330 KKDAIDPLLFYKQVPTKELHESAIVKATQISRRKHLFSRDKLKLFLKQHCPEQGV 389
Db 143 TMYVGVGLL-----RDKTIISTSVKIGEDLIHRRPPVSRNTLKIFRD----- 186

Qy 390 KIKASLSYKIAEDFS--YFPDPT-FTFSPANRRRPPKRIHLSQEDVANKQT 446
Db 187 ---ATSQAPWVHNLAQYGIPIEPNDMMFGGLQKKKK----- 226

Qy 447 LASYSKATKDKLLKQEMKSLAFKAKLREKADALEAKKKEKEDKKEKREBKIV 506
Db 227 -----RREDGPMGDPKKMKNDDEHINVPKYP 255

Qy 507 EERLKKKEERLKVREKEREKUREKRVYLVKQWSPKPMEDCCDLKELPEPTPV 566
Db 256 DD-----LLVQ-----PSADHALLKRP-PL 275

Qy 567 KT--RLPPEIFGDALVLEFLNAGLFDLQDEFDGVTVLEALVGNDSGELCELL 624
Db 276 ATDFRVPKYSGVGLLMVWDFCLSGFGRVNLSP-----FSLVDLENAICHKESNALLVEI- 329

```

```

Qy 625 PFFLTATFOAIABEEVEVAKEQLTDADTKGCSLKSLDLSDCTSLSEILRLHILASGADVTS 684
Db 330 ---HTAIFHLIKDE---GDYFTILRTKRRKL-----VTLVTWAEYLCDFLEMTK 374

Qy 685 ANAKVRYQKRGGFDDTDACMELRLSNPSLVKKLSSTSVYDLTPGB-KMKILHALCGKLL 743
Db 375 TEELTR-----NATVRK-----GYSLIDTDIKUKILRELVEAI 410

Qy 744 TLVSTRDFTEDYDILROAKQFRELKAPQHKEREAAARIRKKEKLEKEQEQMKK 803
Db 411 TTPSVREKLSERVD---QRQALAATKRESTRKAKDEQNSSI-----DGLQDDNESVDE- 460

Qy 804 QEKLEDEORNSTADISIGEEEREDFTSIESKDTQEKLDQDMFTEDDDPGSHKGRGR 863
Db 461 QGKGKEEKDN-----NISRKTEG-----RH 483

Qy 864 GKRGGQGFKEFTRQEQINCVTRELLTADDEEALKQEHQKKEKLEKIOSAIACTNIFPL 923
Db 484 G-----VOHLETEIEKL-SIRSS-----PL 502

Qy 924 GRDRMYRYWIPIPPSIPGLFIEDYSGLTEDMLLPRPSFQNNVQSDPQVSTKTGPEPLMS 983
Db 503 GKDRHYNRYWFPREGRLFVESADS----- 527

Qy 984 ESTSNIDQGRDHSVOLPKPVHKPNRWCFYSSCEOLDQIEALNSRGRHRESALKETLLOE 1043
Db 528 -----KEWGYYSTKEELDVLMSSLVNKLGRALKRQLDKL 563

Qy 1044 KSRICAQLARFSEKFP 1060
Db 564 YSKISNALEKRSKEITH 580

RESULT 31
Q6BV90 PRELIMINARY; PRT; 1230 AA.
AC Q6BV90;
DT 25-OCT-2004 (TReMBLrel. 28, Created)
DT 25-OCT-2004 (TReMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TReMBLrel. 28, Last annotation update)
DE Debaryomyces hansenii chromosome C of strain CBS767 of Debaryomyces
DE hansenii.
GN ORFNames=DEHAOC05082g;
OS Debaryomyces hansenii CBS767.
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Debaryomyces.
OX NCBI_TaxID=284592;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CBS767;
RG Genolevures;
RA Lafontaine I., Sherman D., Fischer G., Durrens P., Casaregola S.,
RA Goffard N., Frangoul L., Aigle M., Anthouard V., Babour A., Barbe V.,
RA Barnay S., Blanchin S., Beckerich J.M., Beyne E., Bleykaeten C.,
RA Boissarie A., Boyer J., Cattolico L., Confanioleri F., de Daruvar A.,
RA Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi A.,
RA Hantraye F., Hennequin C., Jauniaux N., Joyet P., Kachouri R.,
RA Kerrest A., Koszul R., Lemaire M., Lesur I., Ma L., Muller H.,
RA Nicaud J.M., Nikolski M., Oztas S., Ozier-Kalogeropoulos O.,
RA Pellenz S., Potier S., Richard G.F., Straub M.L., Suleau A.,
RA Swennene D., Tekai F., Wesolowski-Louvel M., Westhof E., Wirth B.,
RA Zeniou-Meyer M., Zivanovic I., Bolotin-Fukuhara M., Thierry A.,
RA Bouchier C., Caudron B., Scarpelli C., Gaillardin C., Weissenbach J.,
RA Wincker P., Souciet J.L.;
RT "Genome evolution in yeasts.";
RL Nature 430:35-44(2004).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CBS767;
RA Genoscope;
RL Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; CR382135; CAG85925.1; -.

```


RA Brandon R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle E.G., Helt G., Nelson C.R., Gabor G.L.,
RA Abril J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballou R.M., Basu A., Bakendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhattacharya D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke T.J., Hernandez J.R., Houck J.,
RA de Pablo S., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrieria S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Gary N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hermandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusser D.R., Pacieb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Porci V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Sampson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.Y., Wasserman D.A., Weinstein G.M., Wellenbach J.,
RA Williams S.M., Woodgett, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA Yen R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of *Drosophila melanogaster*";
RN Science 287:2185-2195(2000).
RN [3]
RN
RN SEQUENCE FROM N.A.
RX MEDLINE=22426065; PubMed=12537568;
RA Celniker S.E., Wheeler D.A., Kronmiller B., Carlson J.W., Halpern A.,
RA Patel S., Adams M., Champe M., Dugan S.P., Frislee E., Hodgson A.,
RA George R.A., Hoskins R.A., Laverty T., Muzny D.M., Nelson C.R.,
RA Paciec J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,
RA Svirkas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
RA Weinstein G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.;
RT "Finishing a whole-genome shotgun: Release 3 of the *Drosophila*
RT melanogaster euchromatic genome sequence";
RN Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).
RN [4]
RN SEQUENCE FROM N.A.
RX MEDLINE=22426070; PubMed=12537573;
RA Kaminker J.S., Bergman C.M., Kronmiller B., Carlson J., Svirkas R.,
RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,
RA Ashburner M., Celniker S.E.;
RT "The transposable elements of the *Drosophila melanogaster* euchromatin:
RT a genomics perspective";
RN Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).
RN [5]
RN SEQUENCE FROM N.A.
RX MEDLINE=22426069; PubMed=12537572;
RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA Hradscky P., Huang Y., Kaminker J.S., Millburn G.H., Prochuk S.E.,
RA Smith C.D., Tupy J.L., Whifield E.J., Bayraktaroglu L., Berman B.P.,
RA Bettencourt B.R., Celniker S.E., de Grey A.D., Drysdale R.A.,
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA Lewis S.E.;
RT "Annotation of the *Drosophila melanogaster* euchromatic genome: a
RT systematic review";
RN Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).
RN [6]
RN SEQUENCE FROM N.A.
RP FlyBase;
RG Submitted (SEP-2002) to the EMBL/GenBank/DBSJ databases.
RN [7]

[illegible]

DR PROSITE; PSS0016; ZF PHD_2; 1.
FT NON_TER 796
SQ SEQUENCE 796 AA; 88462 MW; 2EDBBBCAF9720C1E CRC64;

Query Match 4.3%; Score 368.5; DB 2; Length 796;
Best Local Similarity 20.6%; Pred. No. 2.1e-07;
Matches 174; Conservative 104; Mismatches 247; Indels 319; Gaps 30;

QY 793 LKEQEQWKEKQEKL-----KEDEQRNSTADISIGEE-----REDFTSIE 834
DB 4 LRDKVVEGKRLRIIHAHAKYTKGRDTSGIDLGEQHPLGTPTPGKRKKRGSDSYD 63

QY 835 SKTEQKELQDMFTEDDDPGSHKRRGRKGQNGKFETROEQINCVTRELLTAADEEE 894
DB 64 DDDDDSDQGDEDEED---KEDKGGKTD-----ICEDEDE 100

QY 895 ---ALKQEHORKEKELLEKIQS-----AIACNIFPLGRDMRYRYYWIFSIPGLFI 943
DB 101 GDQAASVEELEKQTEKLSKOQSOYRRKLFDAHSLSRWFMFGODRYRRYYILPQCQGI 160

QY 944 E--EDYSGLTEDMLPRPSFQNNVQSDPVSTKTGEPLMSESTSNIDOG-----992
DB 161 EGMESEGLEIAERKELKAESVQIKEMFET-SGDSLNCSNTWDHCQEKDELKENDYT 219

QY 993 -----PRDHSVQLPKP-----1003
DB 220 NLFLOKPGFSKSLKLEVAKMPPSEVMTPKPAGANGCTLSYQNSGXHSLGSVQSTAT 279

QY 1004 ---VHK-----PNWCIFYSCQLDLQILEALNSRHRESAL---KETLL 1041
DB 280 QSNVEKADSNLNFNTGSGFGK--FYSLPLN-DQLLKTLTKRNQWFSLLPRTPCDDTSL 336

QY 1042 QEKSRIACAQLARFSEERFHESDKPOP-----1067
DB 337 THADMSTASLVTSQSPSPSKSPPTAPLGSSAQNPVGLNPPALSPLQVGVSMMGLQF 396

QY 1068 -----DSKPTYSRG-----RSSNAYDPSONCAEQ-----1092
DB 397 CGWPTGVVTNIPTSSVPVLGSLGSLGSEGNSFLTNSVASSKSESPPVONEKATSAQP 456

QY 1093 --LEL-RLRDP-----LLDIERYOQTGLAI-----KV 1118
DB 457 AAVEVAKPVDVFPSPKPIPEEMQFGWMRIIDPED-----LKALLKVLHLRGIREKALQK 510

QY 1119 TDRH--IWSALESGRYEELS-EENKENGIIKTIV--NEDVEMEIDEQKTVVKORLLG 1172
DB 511 IQRHLDYITQACLKNKDVAIELNENEENQVTRDIVENWSVEQAMENDLSVQQVEDLE 570

QY 1173 IKTETPSTVNASTPOSVS-----SVVHYL-----1198
DB 571 RRVASLSIQVGMWCPESAPEREDLVYFEHKSFTFKLCHEHDGETGESAHALERKSD 630

QY 1199 -----AMALFQIOGIERFLK--APLSDASGRSYKTVLDWRBESLLSASLSQVFLH 1250
DB 631 NPDLIAVTRLADLERNIERRIEDIAP-----GLRVRRRALSEARSAAQVALC 678

QY 1251 LSTLDRSVIWKSTILNARKICRKGAENMWLDCGDGRGHHTVCVRPKLTVPYGDFWC 1310
DB 679 IQLQKSIAMEKSIWKVYCIQRGNDEELLCLCDCKGCHTYCHRPKTIITPDGWDFC 738

QY 1311 PECRPKQ-----RCRRLSFRQPSLESDVEDVDSMGGEDDEVGDDEBEGOSEBEVEYEQ 1365
DB 739 PACIAKASGQTLKIKLHVKGKKTNESK-----GKVTTLTGDTDEDSASTSSLKR 791

QY 1366 DEDD 1369
DB 792 GNKD 795

RESULT 36
Q69Z18 PRELIMINARY; PRT; 1369 AA.
ID Q69Z18
AC Q69Z18;

RESULT 36
Q69ZI8
ID Q69Z
AC Q69Z


```
GN Name=Acfl1;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA Nuncio J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,
RA Yu C., Lewis S.E., Rubin G.M., Celniker S.
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY061103; AAL28651.1; -.
DR HSSP; Q9UIG0; 1f62.
DR FlyBase; FBgn0027620; Acfl.
DR GO; GO:0016590; C:ACF complex; IDA.
DR GO; GO:0008623; C:chromatin accessibility complex; IPI.
DR GO; GO:0003677; F:DNA binding; TAS.
DR GO; GO:0042766; P:nucleosome mobilization; IDA.
DR GO; GO:0016584; P:nucleosome spacing; IDA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IDA.
DR InterPro; IPR001487; Bromodomain.
DR InterPro; IPR011011; FYVE PHD_znf.
DR InterPro; IPR001965; Znf PHD.
DR Pfam; PF00439; Bromodomain; 1.
DR Pfam; PF00628; PHD; 2.
DR PRINTS; PR00503; BROMODOMAIN.
DR SMART; SM00297; BROMO; 1.
DR SMART; SM00249; PHD; 2.
DR PROSITE; PS50014; BROMODOMAIN 2; 1.
DR PROSITE; PS01359; ZF_PHD_1; UNKNOWN 1.
SQ SEQUENCE 400 AA; 46108 MW; 84BFE4E8F8FFDCSD CRC64;

Query Match 4.0%; Score 349.5; DB 2; Length 400;
Best Local Similarity 24.3%; Pred. No. 6e-07;
Matches 102; Conservative 81; Mismatches 172; Indels 65; Gaps 11;

Qy 1281 MVLCDGCDRGHTYCVRPKLTVPEGDWFCPEC-----RPQRCRRLSFRQPSLE 1331
Db 1 MLLCDECNAGTHMFLCKPLRSVPPGNWYCNDCVKSLGSLNGQNEKDKKQATKKRKFIV 60

Qy 1332 SDEVDVSMGGEDEVGDDEEGQ-----SSEEEVEVDQEDDSQEEVEV-SLPKRGPPQ 1385
Db 61 EED-----DEATDEEBEKKDDMTDDEAEHENEKHDDEVDDESVTSPSSGRVN 112

Qy 1386 VRLPVKTRKGLSSGFSRSGQQQPGRPYPSRQQSPKTTVSSKTSGRSLRKINSAPPYTK 1445
Db 113 GRILRRPRTRTSRLTSKEIEE-----HAQEDVSDGVSDDA--SLTAGEDIEDSD 164

Qy 1446 SLRIASR-----STRHSHGFLQADVFFVELLS--PRRKRRGRKKSANN 1484
Db 165 EEKVCKQCFYDGGEEKVCQCRFLFHLECVHLKRPRTDFVCKTCKPMPORPRRRHNNNG 224

Qy 1485 TPENSPNPNFRVATKSSGSRSVNIASKLSQESKRRCKRQSPSPFVLGRSS 1544
Db 225 DHDRDEEFP-----KAKRPNRSLRLSIDTARPSNNGNNNNNNNS----SVNNNNHRS 274

Qy 1545 GRCQGVH---ELSAFEQLVVLVRHDDSWPFLKLVSKIQVDDYDIKKPIALNIREKV 1601
Db 275 GRRTNEHPLNSAALYDLLEQIMKHKAAMPFLRPVLTSEVDYDQIIKTPMDLAKIKSL 334

Qy 1602 NKCEYKLASEPIDDIELMFSNCFEYNPRNTSEAKAGTRLQAFHQAOKLGLHVTSPNVD 1661
Db 335 NMGAYQLNEELLSDIQLVFRNCGLYNVEGNEIYDAGCQLERFVIDRCRDMQLPRPSPDMN 394

RESULT 39
Q6CAZ9
ID Q6CAZ9 PRELIMINARY; PRT; 969 AA.
AC Q6CAZ9;
```

```
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Yarrowia lipolytica chromosome C of strain CLIB99 of Yarrowia
DE lipolytica (Fragment).
GN ORFNames=YALI0C23056g;
OS Yarrowia lipolytica CLIB99.
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Dipodascaceae; Yarrowia.
OX NCBI_TaxID=284591;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CLIB99;
RG Genoluvures;
RA Dufon B., Sherman D., Fischer G., Durrens P., Casaregola S.,
RA Lafontaine I., de Montigny J., Marck C., Neuveglise C., Tallia E.,
RA Goffard N., Frangeul L., Aigle M., Anthouard V., Babour A., Barbe V.,
RA Barnay S., Blanchin S., Beckerich J.M., Beyne E., Bleykasten C.,
RA Boisrame A., Boyer J., Cattolico L., Confanioleri F., de Daruvar A.,
RA Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi A.,
RA Hantraye P., Hennequin C., Jauniaux N., Joyet P., Kachouri R.,
RA Kerrest A., Koszul R., Lemaire M., Lesur I., Ma L., Muller H.,
RA Nicaud J.M., Nikoleki M., Oztas S., Ozier-Kalogeropoulos O.,
RA Pellenn S., Potier S., Richard G.F., Straub M.L., Suleau A.,
RA Swennene D., Tekia F., Wesolowski-Louvet M., Westhof E., Wirth B.,
RA Zeniou-Meyer M., Zivanovic I., Bolotin-Fukuhara M., Thierry A.,
RA Bouchier C., Caudron B., Scarpelli C., Gaillardin C., Weissenbach J.,
RA Wincker P., Souciet J.B.;
RT "Genome evolution in yeasts.";
RL Nature 430:35-44 (2004).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CLIB99;
RA Genoscope;
RL Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; CR382129; CAG82483.1; -.
DR InterPro; IPR004022; DDT.
DR Pfam; PF02791; DDT; 1.
DR PROSITE; PS50827; DDT; 1.
FT NON_TER 1
SQ SEQUENCE 969 AA; 112232 MW; 6CA55F880E6029BE CRC64;

Query Match 4.0%; Score 344.5; DB 2; Length 969;
Best Local Similarity 20.7%; Pred. No. 2.6e-06;
Matches 214; Conservative 168; Mismatches 375; Indels 279; Gaps 41;

Qy 153 LLHRKPFVRQKPADLRPDEEVFYCKVTNEIFRHYDDFFERTILCNLSLWCAVTRPGL 212
Db 1 VLFKRAIHFAPIYVVEEDDEVWVIDATGEYFNNDYDLTRMDFYKQKFCIVTGHSL 60

Qy 213 TYQALESE-KKARQNLSPEPLIIPVLYLTSLTHSRLEHICDDIPAYVKDRYFVEET 271
Db 61 TLFEALSKELSGSKSEINDAFENLKEPVLRIQFSTISRLDLVDLTLYGEPKOFFPGEI 120

Qy 272 VEVRNNGARLQCTILEVLPPSHQNGFANGHVNSVDGETIISDSDDSETQSCSPONGKK 331
Db 121 V-----IAIIET-----ATGATERVPA---IIREK--AKFNSIILDSGD- 154

Qy 332 KDAIDPLLFKYQVPTKKELHESAIVKATQISRRRHLFSRDKLKLFLKQHCEPOEGVIKI 391
Db 155 ---VRPATTKYRLEAV-SDPNKVVVCDTQLTRDKNPTKVTLTFTIK----- 198

Qy 392 KASSLSYTKAEQSFSPFDDPTFFTFSP--ANRRGRPPKRIHISQEDNVANKQTLAS 449
Db 199 -----YSAKENW-----NGAPWLVRPEYAEKYRIDQTPVPHLTQ----- 233

Qy 450 YRSKATKRDLLKQEEEMKSLAFELKRAKREKADALEAKKEKEDKEKREE--LKKTIVE 507
Db 234 YRGNPSF--ELRKQRTQDRLKAKQARDNEKNVDIILQOKLKMKENKKNADALIRTLGG 291

Qy 508 EERLKKKEERLKVREKER-----EKLREKRYKRYEVLKOWSKP- 548
Db 292 LSGLKRD---IQVPVSHQNGGCGAGLGAKTALATALASVVTERR-----SPPP 338
```

QY 549 ----REDMECDLKEPPTPVKTRL-PEIFGDALMWLEFLNARGELFDLO-----DRPP 599
 DB 339 KOVIREDEHLRLVGNVPKPVFFHFDKLPDLSLTVSEVLEWTFINVYKEVLVDSFTYDDFI 398
 QY 600 DQWTLF-----VLEELVGNDSGPLCELLFFPLTAIF-----QAIABEE 639
 DB 399 DALRYEGESDLNETHCALLSQVGTNSDSLVE-----YPDIDSDREEDYDEEEEEE 454
 QY 640 EVAKEQITADTK-----GCSLKSLLDSC-----TLSEILRLHI 675
 DB 455 EVKKEEQEVEEVBKEPRGRTRRSQIKKEETEAEABEEBQEEBEEBQEKNSKS 514
 QY 676 LASGADVTSANA-----KYRYQKR-----CG-----PDATD-----ACWE 706
 DB 515 VSPGAENGANYAVFGVDMKTRLQRRNFKDGGWVILGLLYNLTYEPWEDFINGCLD 574
 QY 707 LRLSNPSLVKLSSTSVY-DLTPGKMKILHALCGKLLTVSTRDFIEDYVDILRQAQE 765
 DB 575 VWAAKAPVSLASARDRFDLDTQVKLIQILC-----NLIHPTAIRAYI----- 621
 QY 766 PRELKAQHRKEREAAIRKKEEKLKEQKMKKEQKLEDEQORNSTADISIGBEE 825
 DB 622 -----ECMGRTVLRDKVD-----KQREYKIVSD-----DIRVLENE 655
 QY 826 REDFTSTESKOTEKELDDQDMFTDEDDPGSHKGRGKRGQNGFKFTTQEQINCVR 885
 DB 656 VKRIELEMEEKIPE---DKGFEKKAEGNGDNESEPPSK-----KSKTDNDLSTLR 707
 QY 886 ELLTADEEELKOEHRKEKELL-----EKIQSAIACNTNFPILGRDMRYRYWFP-- 936
 DB 708 TKI-----ODRQETLKEIIAVINAKDEIRFA-DVNRURLLKGORFHRNYWFGN 757
 QY 937 --SIPGLTEEDYSGLTEDMLLPSPSSFQNNVQSDPOVSTKTGSLPMSSESTSNIDQGR 994
 DB 758 GIEDAGGSEEBEGTFTWGLR-----WVGQFEFEIKFLCTGDIPOQKR 801
 QY 995 DHSVOLPKPVKPNRWCFYSCEQLDQIEALNSGRHRESALKETLLQEKSRICQALARF 1054
 DB 802 E--VEGEGVLEDTKWGLHPKELGELIKWLNQWGSREHKLKELLVYLKIKASMYSR 859
 QY 1055 SEEKHFSDKQPOPSK 1070
 DB 860 LEDLGRENELKEDRK 875
 RESULT 40
 ID FUTS_DROME STANDARD; PRT; 5412 AA.
 AC Q9W596; 076891;
 DT 25-OCT-2004 (Rel. 45, Created)
 DT 25-OCT-2004 (Rel. 45, Last sequence update)
 DT 25-JAN-2005 (Rel. 46, Last annotation update)
 DE Microtubule-associated protein futsch.
 GN Name=futsch; ORFNames=CG3064;
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_taxid=7227;
 RN [1]
 RP SEQUENCE FROM N.A.; FUNCTION, SUBCELLULAR LOCATION, TISSUE
 RP SPECIFICITY, AND DEVELOPMENTAL STAGE.
 RC STRAIN=Oregon-R;
 RX PubMed=10839355; DOI=10.1016/S0896-6273(00)81169-1;
 RA Hummel T., Krukkert K., Roos J., Davis G., Klambt C.;
 RT "Drosophila Futsch/22C10 is a MAP1B-like protein required for
 RT dendritic and axonal development.";
 RL Neuron 26:357-370(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RP STRAIN=Berkeley;
 RX MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;

RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Cocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
 RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K.J., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Fostler A., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jaitani M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasako P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kimias I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svrskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Gibbs X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195(2000).
 RN [3]
 RP GENOME REANNOTATION.
 EX MEDLINE=22420609; PubMed=12537572;
 RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
 RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochuk S.E.,
 RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
 RA Bettencourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,
 RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
 RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
 RA Lewis S.E.;
 RT "Annotation of the Drosophila melanogaster euchromatic genome: a
 RT systematic review.";
 RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Oregon-R;
 RX MEDLINE=20196011; PubMed=10731137; DOI=10.1126/science.287.5461.2220;
 RA Benos P.V., Gatt M.K., Ashburner M., Murphy L., Harris D.,
 RA Barrell B.G., Ferraz C., Vidal S., Brun C., Demailles J., Cadieu E.,
 RA Dreano S., Gloux S., Lelaure V., Mottier S., Galibert F., Borkova D.,
 RA Minana B., Kafatos F.C., Louis C., Siden-Kimias I., Bolshakov S.,
 RA Papagannakis G., Spanos L., Cox S., Madueno E., de Pablos B.,
 RA Modolell J., Peter A., Schoettler P., Werner M., Mourikoti F.,
 RA Beilert N., Dowe G., Schaefer U., Jaekle H., Bucheton A.,
 RA Callister D.M., Campbell L.A., Darlamitsou A., Henderson N.S.,
 RA McWilliam P.J., Salles C., Tait E.A., Valenti P., Saunders R.D.C.,
 RA Glover D.M.;
 RT "From sequence to chromosome: the tip of the X chromosome of D.
 RT melanogaster.";
 RL Science 287:2220-2222(2000).
 RN [5]
 RP FUNCTION, SUBCELLULAR LOCATION, AND TISSUE SPECIFICITY.
 RX PubMed=10839356; DOI=10.1016/S0896-6273(00)81170-8;
 RA Roos J., Hummel T., Ng N., Klambt C., Davis G.W.;
 RT "Drosophila Futsch regulates synaptic microtubule organization and is

necessary for synaptic growth." ;
 [6]
 FUNCTION, AND INTERACTION WITH FMR1.
 PubMed11733059; DOI=10.1016/S0092-8674(01)00589-X;
 Zhang Y.Q., Bailey A.M., Matthies H.J.G., Renden R.B., Smith M.A.,
 Speese S.D., Rubin G.M., Broadie K.S.;
 "Drosophila fragile X-related gene regulates the MAP1B homolog Futsch
 to control synaptic structure and function." ;
 Cell 107:591-603(2001).
 CC -!- FUNCTION: During embryogenesis, necessary for dendritic and axonal
 organization and growth at the neuromuscular junction through the
 regulation of the synaptic microtubule cytoskeleton. Microtubule
 hairpin loops are found within a small subset of synaptic boutons
 at the neuromuscular synapse, these loops are stabilized by
 Futsch. Loop morphology and dynamics suggest that rearrangement of
 these microtubule-based loops is a critical component of the
 process of bouton division and for subsequent nerve-terminal
 growth and branching. Translation is repressed by Fmr1.
 CC -!- SUBUNIT: Interacts with Fmr1.
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic. Microtubule-associated.
 CC -!- TISSUE SPECIFICITY: Neuronal cells within the PNS and CNS.
 CC -!- DEVELOPMENTAL STAGE: All stages.
 CC -!- MISCELLANEOUS: 'Futsch' means 'gone' in German.
 CC -!- CAUTION: Ref.4 sequence differs from that shown due to erroneous
 gene model prediction.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.ebi-sib.ch/announce/>
 or send an email to license@ebi-sib.ch).
 CC -----
 DR EMBL; AE003420; CAA45622.3; -;
 DR EMBL; AL031128; CAA20006.1; ALT_SEQ.
 DR PIR; T13564; T13564.
 DR FlyBase; FBgn0015390; futsch.
 DR GO; GO:0005875; C:microtubule associated complex; IDA.
 DR GO; GO:0008017; F:microtubule binding; IDA.
 DR GO; GO:0007409; P:axonogenesis; IMP.
 DR GO; GO:0016358; P:dendrite morphogenesis; IMP.
 DR GO; GO:0007456; P:eye morphogenesis (sensu Drosophila); NAS.
 DR GO; GO:0000226; P:microtubule cytoskeleton organization and b. .; IMP.
 DR GO; GO:0008582; P:regulation of synaptic growth at neuromuscu. .; IMP.
 DR InterPro; IPR0009603; DUF1213.
 DR Pfam; PF06740; DUF1213; 53.
 KW Cytoskeleton; Developmental protein; Microtubule; Repeat.
 FT DOMAIN 1386 3949 53 X approximate repeat.
 FT REPEAT 1386 1419 1.
 FT REPEAT 1430 1456 2.
 FT REPEAT 1539 1566 3.
 FT REPEAT 1577 1603 4.
 FT REPEAT 1607 1635 5.
 FT REPEAT 1672 1699 6.
 FT REPEAT 1707 1735 7.
 FT REPEAT 1754 1782 8.
 FT REPEAT 1791 1819 9.
 FT REPEAT 1828 1856 10.
 FT REPEAT 1865 1893 11.
 FT REPEAT 1902 1930 12.
 FT REPEAT 1939 1967 13.
 FT REPEAT 1976 2004 14.
 FT REPEAT 2013 2041 15.
 FT REPEAT 2050 2078 16.
 FT REPEAT 2087 2115 17.
 FT REPEAT 2132 2160 18.
 FT REPEAT 2179 2209 19.
 FT REPEAT 2272 2308 20.
 FT REPEAT 2620 2643 21.
 FT REPEAT 2678 2704 22.
 FT REPEAT 2737 2763 23.

FT REPEAT	2781	2809	24.
FT REPEAT	2824	2850	25.
FT REPEAT	2873	2904	26.
FT REPEAT	2923	2951	27.
FT REPEAT	2966	2992	28.
FT REPEAT	3006	3034	29.
FT REPEAT	3048	3075	30.
FT REPEAT	3117	3141	31.
FT REPEAT	3145	3173	32.
FT REPEAT	3182	3210	33.
FT REPEAT	3219	3247	34.
FT REPEAT	3256	3284	35.
FT REPEAT	3293	3321	36.
FT REPEAT	3330	3358	37.
FT REPEAT	3367	3395	38.
FT REPEAT	3404	3432	39.
FT REPEAT	3441	3469	40.
FT REPEAT	3478	3506	41.
FT REPEAT	3515	3543	42.
FT REPEAT	3552	3580	43.
FT REPEAT	3589	3617	44.
FT REPEAT	3626	3654	45.
FT REPEAT	3663	3691	46.
FT REPEAT	3700	3728	47.
FT REPEAT	3737	3765	48.
FT REPEAT	3784	3811	49.
FT REPEAT	3812	3838	50.
FT REPEAT	3848	3875	51.
FT REPEAT	3885	3912	52.

Query Match 3.8%; Score 329.5; DB 1; Length 5412;
 Best Local Similarity 19.4%; Pred. No. 7.2e-05;
 Matches 333; Conservative 254; Mismatches 643; Indels 487; Gaps 74;

Qy	4	ASESSR--GVAPLNNVVLPGSLPLPSVVT--GCKSHR--VANKKVEA-----	46
Db	3213	SQEAPESEAESESLKDAAPSQETSRPESVTEYKDGKSPVASKEASRPASVAENAKDSA	3272
Qy	47	RSEKLLPTALPPSEPK--VDQKLPRSS-----ERRGGGTGTPPARSRVAAGEA	94
Db	3273	DESKEQPESLPQSKAGSIKDEKSPASKDEAEKSESRRESVAEOPPLYSKEVSRPAS	3332
Qy	95	AARGAAGPERGSPGLGRVSRCLSCGEGQVAVGVIAK---RGRRGDSRRRAPGREGM	151
Db	3333	VAESVKDEAEKS---KEESP--LMSKEASRP--SVAGSVKDEAEKSESRRESVAEKS	3385
Qy	152	PLLRKPFVQKPPADLRDPDEEVFYCKVTNEIFRYDDFFERTILCNSLWSCAVTGRPG	211
Db	3386	PL-----PSKEASRPASVAESVK-----	3403
Qy	212	LTQEALESSEKKARQNLQSPPEPLIIPVLVLTSLTHRSRLHEICDDIFAYVKDRYFVEET	271
Db	3404	---DEADKSEESRRSGAEKSP-----ASKEASRPASVAESI---KDE--AEKS	3446
Qy	272	VEVIRNNGARLQCTILEVLP--PSHQNGFANGHVNSVDGETIIISDSDSETQSCSFQNGK	330
Db	3447	KEESRRE-----SVAEKSPLPSKEASRPTSVAESVKDE-----AEKSEES	3487
Qy	331	KQDAI---DPLLFKYQVQPTK--KELHESAIVKATQISRRKHLFRDRLKLFKQHCERQOE	386
Db	3488	RRDSVAEKSPLASKEASRPASVAESVQDEAEKSESRRESV--AEKSPLASKEASRPAS	3545
Qy	387	GVIKIKASSLSLYK-----IAEQDFSYFFDDPTTFIFSP--ANRRGRP-----	429
Db	3546	VAESIKDEAEKSESRRESVAEK-----SPLASKEASRPTSVAESVKDE	3590
Qy	430	-PKRIHISQEDNVANKOTLASR-----SKATKERDKLLKQBEKMSLAFKAKL--KRE	480
Db	3591	AEKSEESRDSVAEKSPLASKEASRPASVAESVQDEAEKSESRRESVAEKSPLASKE	3650
Qy	481	KADALEAKKKEKEDKKEKREELKK--IVEERLKKKEK-----ERLKVREKEREKL	532
Db	3651	ASRPASVAESVKDDAEKSESRRESVAEKSPLASKEASRPASVAESVKDEAEKSESR	3710

Db	592	RKLVDVS	-----KPKEDNSALE-----KTVQDEEATVKEEDRSARAEPLE	631
Qy	774	-----HRKEEBEAAAARTRKKEEKLEKQEQKQKQKLEKDEBORN-----STADISIGE	823	
Db	632	NTNLSVCREGQETEVKDVVKNELFKD-----KAGPDNRVKGKSAADL-----Q	676	
Qy	824	EEREDFDTSIESKDTQEQKELQDDMETEDDOPGSHKR-GRGKRQNGKFKEFTTQEQINC	882	
Db	677	BEAEDY-----KKSFAEVVVERSEGEGRKRDGIKEKEEDGKTTSEODE---	722	
Qy	883	VTRELLTADBEALKQEHQKQE-----KELLEKIQSAIACTNIPPLGRDMRYRYWTFP	936	
Db	723	---DLLKDGKVGVEQHDSDGKTKTEETDPKQSD-----CKEXTEE-----P	765	
Qy	937	SIPGLFIEDYSGLTEDMLLRPSSFQNNVQSDPQVSTKGTGEPLMSESTSNIDGPRDH	996	
Db	766	KEEGSKVEEDVTA-----DRSERKEMEELNEEGAKRETETESTKRDKEEKED	811	
Qy	997	SVQLPKPVHKPNRMCFYSSCRLDQLIEALNSRGHRESALKETLLQEKSRICQAQLARFSE	1056	
Db	812	VTKNDSDVEKCK-----TEAEAAVQPLR-KGSRPLRPRGKPKQEBE-----PE	853	
Qy	1057	EKFHFDKPDQDSKFTYGRGRSSNAYDPSOMCAEKQLELRRLRDLFDIEDRIYQGTLGAI	1116	
Db	854	TK-----KPKPAPK-----KRGK-----PKKKVEVSSODEKTSQDEGFSQ-----	888	
Qy	1117	KVTRDHIWRSALSGRYELLSEENKENGIIKTVNEDVEEMEIDEO-----	1161	
Db	889	---EDBE-----EKESKKGVSKVEVMEDEDEDEDEDEDEDEDEDEDDDDDDDDDD	932	
Qy	1162	---TKVIVKDRLLGIKT-----ETPSIVSTNASTPOSVSVVHYLAWALFOIQEG	1208	
Db	933	YGKKKIQKKK---GVNTAKKKKKKKRLRNTPK-----RKASQKAVNKIVE-----MEKDEEG	981	
Qy	1209	IERFLKAPLDASDGSGRYKTVLDRWRRESLLSASLSQVFLHLSTLDRSVIWSKILNAR	1268	
Db	982	GDAR-----KNSYAGKN-----KXKADSDSWISENDTTP	1012	
Qy	1269	CKIKRKGDAENMLCDGCDRGHHTYCVRPKLKTVPEGDFWFCPCRCRQKRRRLSFRORP	1328	
Db	1013	CKCKGLYNHPRWILLCDKCDSGFHTACLRPLMAIPDGNWFCPCKEHEBELIVNLOAKLE-	1071	
Qy	1329	SLESDVEDVDSNGGED-----DEVDGDEEGQSEEBEYVE	1364	
Db	1072	DLDGDKKKRDIARDERKLYKYTDVCAVNILPDTWVIEKKKQKQKRRRNDYBEE	1131	
Qy	1365	QDEDDSQE-EBEVSLPKRGRPOVRLPVKTRGKLSFSFSSRGOQOQPGRYPSRQOSTPKT	1423	
Db	1132	EDEEEVEYEDQRNHSSRSR-----RSHGQSQRSSQSGRRSSQSRHRRDRSRPRR	1184	
Qy	1424	T-----VSKYTGSRSLRKI-----NSAPPT-----ETKSLRIASSTRHS	1457	
Db	1185	SORDVWPDPFEDYEWVWRRSQRSRKTIIDYSPKDPFDSAIKTAICDEVBEHKKAVDQIKSR	1244	
Qy	1458	HGPLQADVVELL-----SPRKRGK-----SANTT	1485	
Db	1245	WGVSRGKDMSTIMGDEEQQEKEEKAEEVDGAEEGKRRDRKKGKRRRLNDLDGSSE	1304	
Qy	1486	PENSPNFPNFRVIATKSSEQRSVNIASKLSIQESE-----SKRRCKRQSPSPSVT---	1538	
Db	1305	DENSEDF-----KVDSENESESADGNSFGASDSDSLTSRRRASRRKASKRSLRYAG	1358	
Qy	1539	LGRSSGR	1546	
Db	1359	MPRRKSR	1366	
RESULT 42				
ID	Q9NYU0	PRELIMINARY; PRT; 1189 AA.		
AC	Q9NYU0;			
DT	01-OCT-2000	(TrEMBLrel. 15, Created)		

DT	01-MAR-2001	(TrEMBLrel. 16, Last sequence update)
DT	01-MAR-2004	(TrEMBLrel. 26, Last annotation update)
DE	HBV pX associated protein-8.	
OS	Homo sapiens (Human).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
OX	NCBI_TaxID=9606;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RX	MEDLINE=21895855; PubMed=11788598; DOI=10.1074/jbc.M11354200;	
RA	Shamay M., Barak O., Doitsh G., Ben-Dor I., Shaul Y.;	
RT	"Hepatitis B virus pX interacts with HBxAP, a PHD finger protein to	
RT	coactivate transcription."	
RL	J. Biol. Chem. 277:9982-9988 (2002).	
DR	EMBL; AF227948; AAF61709.2; --	
DR	HSSP; Q9UIG0; IF62.	
DR	GO; GO:0003677; P:DNA binding; IEA.	
DR	GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.	
DR	InterPro; IPR001965; Znf_PHD.	
DR	Pfam; PF00628; PHD; 1.	
DR	SMART; SM00249; PHD; 1.	
DR	PROSITE; PS01359; ZF_PHD_1; UNKNOWN_1.	
DR	PROSITE; PS00016; ZF_PHD_2; 1.	
SQ	SEQUENCE 1189 AA; 134958 MW; B5A24443F3C6B259D CRC64;	
Query Match 3.8%; Score 326; DB 2; Length 1189;		
Best Local Similarity 18.5%; Pred. No. 1.9e-05;		
Matches 254; Conservative 177; Mismatches 455; Indels 484; Gaps 47;		
Qy	462	LKQEM-----KSLA--FEKAKLKREKADALEAKKKE---KEDKEKREELKKIVBEER 510
Db	1	MKSEEQMDLENRRSTANVLEETTVKKEKEDEKELVKLPVIVKLEKPLPENBEKKIIEES 60
Qy	511	LKKKEERLKV-----REKER-----E 529
Db	61	DSFKENVPKIVKEVKECRADPKDTKSSMEKFPVQEPERIEFGNIKSSHEITEKSTEETE 120
Qy	530	KLREKKRYVEYLQWQSKPRDME-----CDDLKELPEPTPVKTRLPPEIFGDALMVL 582
Db	121	KLKNDQAKIPLKREIKLSDDFSPVKGPICKSV-----TPTKEFLKDELKQSE---- 170
Qy	583	EFLNAGELFDLQDFPDGVTLEVEALVGNDSGLCELLFFFLTAIFAQIAEBEEV 642
Db	171	ETCKRISITALGHEGKQLVGEVSDERVAPNFKTEPI-ETKFY-----ETKEESY 220
Qy	643	AKEQLTDADTKGCSLKSLD--LDCTLSEILR-LHILASGADVTSANAKYRKQKGGPDA 699
Db	221	SPSKDRNIITEGNGTSELSNVITSMKTGELEKETAPLRKQAD-SSISVLTHSQAQIEE 279
Qy	700	TDDACMELRLNPSLAVKLSSTVYDLTPGEMKILHALCGKLLTLVSTRDF---IEDTV 756
Db	280	PDPPEMTSLDSSEWAKDLSSKTLSSSTESCTMK-----GEEKSPKTKDKRPPILECL 333
Qy	757	DILQAKQEFRELKAEQHRKEEAAAIRKRKEELKEQEQKMKKEQ-----EKL 807
Db	334	EKLESKKTFDKDAQRLSPIEE---VPKSTLESEKPGSPAAETSPSPSNIIDHCEKL 389
Qy	808	---KEDEORNSTADISIGEEREDPDTIESKDTQEKELD--QDMFTDEDDPGSHKGR 862
Db	390	ASEKEVVEQSTSTVGGOSVKVDLETLKEDSEFTKVMENLDNNAQTSIGIEESETK--- 446
Qy	863	RGKRQNGKFKEFTTQEQINCVTRELLTADBEALKQEHQKKELEKLEKIQSAIACTNIPP 922
Db	447	-GSMQSKFK-----YKLVPEETTASENTTETSERQKEGIKLITRISS----- 489
Qy	923	LGRDMYRRYWIFFSIPGLFIEEDYSGLTEDMLLRPSSFQNNVQSDPQVSTKGTGEPLM 982
Db	490	--RKKK-----PDSPPKVL-----PENKQEKTEKEEK---TNVGRTLR 524
Qy	983	SESTSNIDGPRDHSVQLPKPVHKPNRMCFYSSCRLDQLIEALNSRGHRESALKETLLQ 1042
Db	525	-----RSPRISRPTAK-----VAEIRDQKADKKRGEDEVEEESTALQ 563


```
QY 1043 EKSRIQAQLARSEKHFSDKPPDQSKFTYSGRSSNAYDPSQCAEKQLELRDLRDFLL 1102
Db 564 K-----TDKGIKXSEKDTNSKVS-----584
QY 1103 DIEDRIYQOTLGAIVTDRIHWSALESGRYELLS- EENKENGIIKTWNDEVEEMEIDBQ 1161
Db 585 -----VPGKGVRTGRTGRWKYSNDESESGSEKSAASEEKEKESE 631
QY 1162 TKVIIVKRLGLIKTETPTSTVSTNASTPOSVSVVHYLAWALFQIEQGIERRFLKAPLDAS 1221
Db 632 BAILADD-----638
QY 1222 DSGRSYKTVLDWRBSLSSASLSQVFLHLSTLDRSVIWSKILNARCKICRKGDAENM 1281
Db 639 -----DEPCKKGLPNHPELI 654
QY 1282 VLDGCDGRGHHTYCVRPKLKTVPEGDWFCPECRKQRCRL-----SPROPSPLES 1332
Db 655 LLCDSGDSGYHTACLRPPLMIIPDGEWFCPCOHKLLCEKEEQLOLDLVALKKXERABR 714
QY 1333 DED-----VEDSMGGEDDEVGDDEEGQSEEEYEV-----1363
Db 715 RKERLVVYVGISIENIIPQEPDFSEDQEKKKDKSKKANLLERRSTRTRKICISYRDEF 774
QY 1364 -----EQEDDSQE-----EEVSLPKRGRPQVRLPVK 1391
Db 775 DEADAIEDDITKEADGGGVGRGKDITITGHRGKDITILDDEERKENKRQPORAAAARRK 834
QY 1392 TRGLSSSFS-----SRGQQQ-----PGRVPSRSQSTPKTTVSSK--TGR 1431
Db 835 KRRRLNDLSDNLDEESEDEBFKISDGSQDFVVDENPDDESEDDPSNEDSDTDFCSR 894
QY 1432 SLRKINSAPPTETKSLRIASRSTRSHGFLQ-----ADVFVLLSPR 1473
Db 895 RLRRHPSRPMQSRRLRKTPKKYSDDDDEEESSENSRDSFSDDFDVFV-TRRR 953
QY 1474 RKRRGRKANNTPENPNPFRVATKSSSEGRSVNTASK--LSLQES-----1520
Db 954 RSRNRQKQINYNKEDSESDG-----SQSLRKGKEIRRVHKKRLSSSESESYLSKNSD 1008
QY 1521 -----ESKRRCKR-----QSPEPSPVTIGRESSGRQGVHELSAFEQLVVELVRHD 1567
Db 1009 DELAKESKRVKRGSRSTDEYSEADEEEBEGKPSRR--LHRIETDEESCSNNAHGD 1065
QY 1568 DSWPFLKLVSKIQVDDYDIINKPIALNIIRKVNKCRYKLASEFIDIE 1617
Db 1066 ANQPARDQPRV-LPSEQESTKKP-----YRIESDEEDFE 1100

RESULT 43
Q81SF6 PRELIMINARY; PRT; 18519 AA.
AC Q81SF6;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE 2Ma 2 protein.
GN Name=isoF;
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI TaxID=6239;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=22269627; PubMed=12381307; DOI=10.1016/S0022-2836(02)00970-1;
RA Flaherty D., Gernert K., Smeleva N., Tang X., Mercer K.,
RA Borodovsky M., Benian G.;
RT "Titins in Caenorhabditis elegans with Unusual Features: Coiled-coil
RT Domains, Novel Regulation of Kinase Activity and Two New Possible
RT Elastic Regions.";
RL J. Mol. Biol. 323:533-549(2002).
CC -i- SIMILARITY: Belongs to the Ser/Thr protein kinase family.
DR EMBL; AY130759; AAN61518.1; -.
```

```
DR PIR; E89066; E89066.
DR PIR; T33247; T33247.
DR HSSP; Q10466; 1BEV.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0004674; F:protein serine/threonine kinase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0006468; F:protein amino acid phosphorylation; IEA.
DR InterPro; IPR008938; ARM.
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR008957; FN_III-like.
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003598; IG_c2.
DR InterPro; IPR011009; Kinase_like.
DR InterPro; IPR006025; Pept_M_Zn_BS.
DR InterPro; IPR000719; Prot_Kinase.
DR InterPro; IPR002290; Ser_thr_pkinase.
DR InterPro; IPR008271; Ser_thr_pkin_AS.
DR Pfam; PF00041; fn3; 11.
DR Pfam; PF00047; ig; 6.
DR Pfam; PF00069; Pkinase; 1.
DR SMART; SM00060; FN3; 11.
DR SMART; SM00408; IGC2; 16.
DR SMART; SM00220; S_TKc; 1.
DR PROSITE; PS50853; FN3; 11.
DR PROSITE; PS50835; IG_Like; 38.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN 1.
KW ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.
SQ SEQUENCE 18519 AA; 2051739 MW; 6A8441C5D0BA7729 CRC64;

Query Match 3.7%; Score 320; DB 2; Length 18519;
Best Local Similarity 19.6%; Pred. No. 0.00069;
Matches 344; Conservative 249; Mismatches 610; Indels 548; Gaps 74;

QY 206 VTGRPLTYQBALSEKKAQNLOS-----FPEPLIPVLYLTSLTHRSRL 251
Db 12672 VTGKP-LEAKPVEDKQDASQPSSESSPPTDGKKKQIPKALFIP-----12717
QY 252 HEICDDIYAYVKDRYFVEETVEIRNNGARLQCTILEVLPSSHONGFAN-----GHVNS 305
Db 12718 ---DEISSRRGD-----PSTHSETNITTIRREGS 12746
QY 306 VDGETIILSDSDS-----ETQSGSFQ-----NGKKDAIDPLLPKYK-----343
Db 12747 ADATPLVEPLSASVMKVESAKAEFSKRSETPDDKSRKKEGLPPAKSEKDEV 12806
QY 344 -----VOPTKKEHESAI-----VKATQISRKHLSFRDKL--K 375
Db 12807 AEKQSTEALIESKKKEVDESKISEQQPSDKNKKSEVVGVPEKAAGPETKKDVSEIEVPKK 12866
QY 376 LFLKHQCEPQSGVIKASSISTYKIAEQDSYFPFDDPPTFIFSPANRRRGRPPKRIHI 435
Db 12867 KTIKKTEKSDSSISQKSNVL---KPADD-----K 12894
QY 436 SOEDNVANK-----QTLASVRSKATKERDKLLKQEEWKSIAFEKA-----KLKREKAD 483
Db 12895 SKSDVDTDKSKTTEDQTKVATDSKLEKAD--TTQIETETVDDKSKKVLKKTKEKD 12953
QY 484 ALEAKKKEK-----EDKEKKREELKXIVVEERLKK-KEEKERLKVREKEREKLEKRY 538
Db 12954 SFISQKSETPPVPEPTKPAESEAQKIAEVNKAQKQKVEDDNLKRAEVAAKKIADEKLKI 13013
QY 539 -----VEYLKQWSKPFREDMECD-----DLKELPEPTPVKTLPEIFD 577
Db 13014 EAEANIKKTAEEVAEAKQKEDQKLETEVVVSKSAAEKLEKQAQIKKAAE-----AD 13069
QY 578 ALMWLEFLNAGFELFDLQDEFPDGVTVLVEALVGNDSGLPCCELLPFLFLTAIFQAI 637
Db 13070 AVKKQELNEKNKLEAAKGAADKUKLSESAKSKKVSSESV-----KFG 13116
QY 638 EEEVAKELQTDADTKGCSLKSLD-----LDSCTLSELRLHLILASGADVTSANA- 687
```


Db 12672 VTGKP-LEAKPVEDKKDASQSSKSSPPDGGKKKQIPKALFIP----- 12717
QY 252 HEICDDIFAYVDRYFVEETVIRNNGARLOCTILEVLPPSHONGFAN-----GHVNS 305
Db 12718 ----DEISSRFGD-----PSTMHSETNITTTIRREGS 12746
QY 306 VDEGIIIIISDSDS-----ETQSCSFQ-----NGKKDAIDPLLFYK----- 343
Db 12747 ADAKTPLVPLSASVMKVESAKEABFSFRSRSETPDDKSRKKEGLPPAKKSEKDEVT 12806
QY 344 ----VQPTKXELHBSAI-----VKATQISRKHLSRDLK-K 375
Db 12807 AEKQSTEALIESKKEVEDESKISEQOPSDKNKSEVVGVEKAAGPETKQDVSEIEBVPK 12866
QY 376 LFLKQHCPEQGVIIKASSLSTYKIAEQDFSYFFPDDPPTTFISPANRRRGRPPRIHI 435
Db 12867 KTIKKTEKSDSSISQSNVL---KPADD-----K 12894
QY 436 SOEDNVANK-----QTLASYSRKAATKERDKLLQBEKMSLAFKA-----KLKREKAD 483
Db 12895 SKSDVDVTKSKTTEBDQTKVATDSKLEAAD-TTKQIETETVDDKSKKKVLLKKTEKSD 12953
QY 484 ALKAAKKEK-----EKEKKEKREKKTIVBERLKK-KEEKERLKVREKEREKLEBEKRY 538
Db 12954 SFISOKSETPPVVEPTKPAESAQAIAEVNKAQKEVDNULKRAEVAAKKIADEKUKI 13013
QY 539 -----VEYLKQMSKPREDMCD-----DLKELPEPTPVKTRLPPEIFGD 577
Db 13014 BAEANIKKTAEEAAKQKEDEQLEKLETEVVVSKSAEKLKLEKQAIKAAE-----AD 13069
QY 578 ALMVLEFLNAGFELFDLQDEPDGVTLEVLBALVGNDSGLCELLFFFLTAIFAIAE 637
Db 13070 AVYKQKELNKNKLEAAKSAADKLEBESAAKSKVSESV-----KFG 13116
QY 638 BEEVAKQELTDADTKGSKSLD-----LDSCTSELTLHLILASGADVTSANA- 687
Db 13117 EKTATAGETVQVESEPTSKTIDTKVKGATEPADETPKKIITKKTEKSDSSISQKSA 13176
QY 688 ---KYRQKRGGFADTDACMELRL-SNPSLVKLSSTSVYDLTPGERKMTILHALCGKLL 743
Db 13177 DSEKYSKQKEQD-EPTKPAVSETQVMTREADSKQKETD-----EKLDAEIAAK-- 13226
QY 744 TLVSTRDFIEDVVOILROAKQFRELKA-EQHRKEREBAARIRKKEEKLK-EQE---Q 798
Db 13227 -----TKQEADEKSLDAQEKIKVYSEDDAARKEKELNDKLESEIATK 13271
QY 799 KMKEKQEKLEDEQRNSTADISIGEEEREDFDTSESKDTQEKELDQDMFTEDDQPSH 858
Db 13272 KASADKLEQQAQKAAEVEAAKQKE-----KD-EQLKLDTE-----AAS 13313
QY 859 KGRGRGKRGNGFEFTREQINCVTRELLTADBEALKQEHQKEKELLEKIQSAIACT 918
Db 13314 KKAEEK-----LELEKQAQIK-----KAAGADAVKQKELDEKKNLEANKSAA-- 13358
QY 919 NIFPLGRDMRYRWYIFPSIPGLFTIEBYSGLUTEDMLLPRPSSFQNNVQSQ----DPQVST 975
Db 13359 ----GK-----LKTEERSA-----AKSQTVQEEQAKLDAQTKA 13387
QY 976 KTGEPIM-----SESTNSNDQGRHDSVOLPKPVHKPNRWCYFSSCEQLDLIE 1024
Db 13388 KTAEQTKLEKDEKSTKSEKSETVDEKPK--KKVLKXKTEKSD-----SSISQKSETSK 13440
QY 1025 -ALNSRGHRES-----ALKETLLQKSRIACAQLA-RFEEKGFHFSDKPPQPSKPTY 1073
Db 13441 TVVESAGSESESTQKVAADARKQKTEDEKQKLEAIEITAKSNADEK-----SKLEAESKL-- 13494
QY 1074 SGRSSNAYDPSQMCBAKQELRLRDLFLDIEDRIYQGTGLKAVTDR-HIWRSALESG- 1131
Db 13495 ---KKAEEVAAKQKQKDEQK-----LDTEAASKKAAAEKLEKQSHIKKAAEVDVAV 13546
QY 1132 -RYELLSENKENGLIKTVNEDVEIMEIDEQTKVIVKDRLLGI----- 1173
Db 13547 KQKELEBKQLESEAAATKKAADAEKLEQKKAEEIALIEIQKEQEKLAQEQSRLEDE 13606

QY 1174 -----KTEPSTVSTNASTPQSVSSVVHYLAMALFQIEQGIERRFLKAPLADS-- 1223
Db 13607 AKKSAEKQKLESETSKQTEEAPKE-----SVDEKPKKKVLKKKTEKSDSI 13653
QY 1224 ---GRSYKTVDLRWRESLLSSASLSQVFLHLSTLDRSVIWSKSIINARCKICRKGDAEN 1280
Db 13654 SQKSSAKSTVDA-AETLESDFNL-----VEKTVQKEQSPDESTATIKRDPQAQKTEBI 13708
QY 1281 MVLCDGCRGHHTYCVRPKLTVPBGDMFCRQCRKQCRRLSRQRPSLESD----- 1333
Db 13709 SKQDDGDEKKTTTIDGKPPK-----PEDS-----EATPKRVVKKTKQKSDSVASDASLAVS 13760
QY 1334 ---EDVE-----DSMGGBDEVDGDEBEG---QSEBEEVEYEQDE---DD 1369
Db 13761 KLSDDVEBKPKKKVLKKKTEKSDSVISETSSVDTIKPSVEIPTKEAQEMILHNRFSTDS 13820
QY 1370 SOEBEVELPKRGRPOVRLPVKTRGKLSRSSSSRQOQEPGRYPSPRSQOSTPKTTVSSKT 1429
Db 13821 AVESPEKNAHKDDTEKTTDDMTTRKSAIFSDDEQ-----SISST 13862
QY 1430 G-----RSURKINSAPPTETKSLRIASRSTRSHSGPLQADVVELLSPRKRRGRKSANNT 1485
Db 13863 SSEGRRRRRTGFASKFASDTLAL-----RGDNVEIEAELLAEDDTVTWKVNGKQADLNS 13917
QY 1486 PENSFPNFPNFRVIATKSSEQSRS-VNIASKLSLQSESKRRCRKRQSPSPSPVT---LG 1540
Db 13918 RCHMSHTFRTLIIDEVPTDSGMEITATCG---TESHTTILK---VEELPVDPVKVL 13971
QY 1541 RRSGRQG-----GVEHLSAFQVLVLEVRHD 1567
Db 13972 RKTSGKEQEVTVISVTLNHPIDISKVWVLKDGKPLEINKDYSIDTVGCVSVLTLRAKYE 14031
QY 1568 DSWPFLKVLVKIQPDYDIKKPIALNIIEK---VNK-CEYKLASEFIDDIEMPS- 1621
Db 14032 DSGKYVVDGVDGDCSTHLSIOGKPVLRKNSVETKPVITVDKDDQFELLVAYDSNPEASFSM 14091
QY 1622 -----NCFEYNPRNTSEAKAG---TRLQAFPHIOAQKLGHLV--TP 1657
Db 14092 TVDGDKLEDFGRSRIIDVVDGKLTGKRGVSKTDAGEYEVKLKNFGEVAQKFDVKVNDTP 14151
QY 1658 SNVDQVSTPPA 1668
Db 14152 SAPGDVSVVKA 14162

RESULT 45

Q06166
ID Q06166 PRELIMINARY; PRT; 1661 AA.
AC Q06166;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1998 (TReMBLrel. 08, Last sequence update)
DT 01-MAR-2004 (TReMBLrel. 26, Last annotation update)
DE Mature parasite-infected erythrocyte surface antigen.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Palo Alto;
RX MEDLINE=92158014; PubMed=1741020; DOI=10.1016/0166-6851(92)90231-8;
RA Coppel R.L.;
RT "Repeat structures in a Plasmodium falciparum protein (MESA) that
RT binds human erythrocyte protein 4.1";
RL Mol. Biochem. Parasitol. 50:335-347(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Palo Alto;
RX MEDLINE=99172328; PubMed=1007328; DOI=10.1006/expr.1998.4374;
RA Kun J.P., Waller K.L., Coppel R.L.;
RT "Plasmodium falciparum: structural and functional domains of the
RT mature-parasite-infected erythrocyte surface antigen";
RL Exp. Parasitol. 91:258-267(1999).


```
Db 547 -----TKSVBDEVDLSKVPTKEKEDMAEQOQCHDSQKARGNGCAETEERAEDTVL 601
QY 795 EOEQMKKEQKEDKEDEORNSTADISIEGEREDPDTIESKDTQEKELQDMFTEDEDD 854
Db 602 SSKRSPAEKSAKAKNDSSAAA-----ATDLPVALVKPKVEK----- 639
QY 855 PGSKRGRGRKGQNGKFEFTRQEOINCVTRELLTADDEEALQKHORKEKELLEKIQSA 914
Db 640 -----RRGRKSNTPKK-AFDSSVTASTHQMTEDEEDTKG---VKVEAAVQASGA 688
QY 915 IACTN-----IFPLGRDMRYRWPSPISGPIFIEDYSGLTEMLPRPSSFNQVOSQD 970
Db 689 AGAKNGDHTVAPSSEK-----PKKELKE-----SDVLSRET-----E 723
QY 971 POVSTKTEPLMSESTNSIDGPRDHSVQLPKPVHKPNRWCFYSCEQDLQILIALNSRG 1030
Db 724 KQVSATAG-----LHDKKGAKAIPKDVVK-----TEPLKEED 755
QY 1031 HRE-SALKETLLQEKSRICAOALARFSEKPFHSPDKPOPDSPKTYSRGSSNAYPSQWCA 1089
Db 756 KQEQAEDKQAVESTDASMLKPSKEKTDEASKAATDVKEINVQPDPPATEPDQGES 815
QY 1090 EKQELRLRDL-----LDIEDRIYQGTGLGAIKVTDRHWRSALESGRYELLSEENKENGII 1146
Db 816 PPKRRHRHNGLVGLDLSMVLDAAGSGGNNPPVRQSRRIAMQK----- 859
QY 1147 KTVNEDVEEMIDEQ-----TKVIVKRLILGK-----TETPSTVSTNASTPOSVS 1192
Db 860 --IKEETNRREIEDQMLKMKADAVKKKKGLMKLSDDDYRADRTSPDHSSDSSGGV- 916
QY 1193 SVVHYLAMALQIEQGIERRFLKAPLDASDGRSVKTVLDRWRRESLSSASLSQVFLHLS 1252
Db 917 -----GADRRRKKKKKKKSLQOQOQAAQASTWNSAPSE----- 955
QY 1253 TLDRSVIWKSILNARCKICRKGDAENMLCDGCDRGHTTYCVRPKLKTVPEGDFWCPPE 1312
Db 956 -----SSSETEAEDYFAEPYHSEED-----GYHCACLKPVLFITPEGDFWCPV 998
QY 1313 CRPKQRCRLFRQRPSPLESDEVEDSDMGDEDEVDGDEEGQSEEBEYVEQEDDSQOE 1372
Db 999 CLHRQLIER-----LQSKLEYDALRAKLEAETKRLLEEASARAELAEQALANER 1050
QY 1373 EEEVSLPKRGPVRLPVKTKGLSSSPSSRGQOQEGRYPSRQOOSPPTKTVSSKTGRS 1432
Db 1051 EEQRAQRAQQOQL--LEADABGRSGSGSEESSRRRRRRQRHRRRAPKRRSSR 1108
QY 1433 LRKINSAPPTTKSLRIASRSTRHSHG---PLQADVVELLSPRKRRGRKSA-NNTPEN 1488
Db 1109 SRLDSALRMRWGCEIVEEBEDEDGFGSKGKDISTIIIEATTREEQARKKALQKODEE 1168
QY 1489 SPNFPNFRVIATKSE-----QSRVNIASKLSLQESKRRKRQSPSPVTLGRR-S 1543
Db 1169 EAEGEERLNGTOEQDQGEHGKTARAEKG-SEDEDEGKGAVGQRKAPTASGRAVK 1227
QY 1544 SGRQGG 1549
Db 1228 KSKSGG 1233
```

RESULT 48

```
Q9FND5
ID Q9FND5 PRELIMINARY; PRT: 2910 AA.
AC Q9FND5;
DT 01-MAR-2001 (TRENBLrel. 16, Created)
DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)
DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)
DE Similarity to heat shock protein.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
```

```

[1]
RN SEQUENCE FROM N.A.
RP MEDLINE=98069011; PubMed=9405937;
RA Kotani H., Nakamura Y., Sato S., Kaneko T., Asamizu E., Miyajima N.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. II.
RT sequence features of the regions of 1,044,062 bp covered by thirteen
RT physically assigned P1 clones.";
EL DNA Res. 4:291-300(1997).
DR EMBL; AB006702; BABI1602.1; -.
DR EMBL; AB017062; BABI1602.1; JOINED.
DR GO; GO:0006457; P:protein folding; IEA.
DR GO; GO:0006986; P:response to unfolded protein; IEA.
KW Heat shock.
SQ SEQUENCE 2910 AA; 325353 MW; A847EC3FE1427DF7 CRC64;

Query Match 3.7%; Score 315.5; DB 2; Length 2910;
Best Local Similarity 19.3%; Pred. No. 0.00014;
Matches 313; Conservative 263; Mismatches 616; Indels 431; Gaps 76;

QY 160 VRQKP---PADLRDDEEVFYCKVTNEIFRHYDDPFERTILCNSLVMSCAVTGR---PGL 212
Db 625 IQERPSSKASLEPKEEVDH--ISNTEEH-EHVLERDVQOCETIESEAVETKEDTOPSL 681
QY 213 TYQALESEK-KARONLQSPPEPLIIPVLYLTSLTHSRSLHEICDDIFAYVKDR-YFVEE 270
Db 682 DLKEDKETEAEPTKTVFSSDE-----VRSSAVQEQFGEHTEPCSSBIKDESHGKEE 734
QY 271 TVEV-----IRNGARLOCTILEVLPPSHONGFANGHNSVDGTIIISDDSETQSCS 325
Db 735 SVEVKSQETVQDENTEDKHDVLKV--PSTESEKTYQGN---EPETVLVSNSTGYEK--- 784
QY 326 FQNGKKDAIDPLFKYKVOPTKELHESAI---VKATOISRRK-HLFSRDLKLFLKQ 380
Db 785 ----SEKSPSDLVL-----NVDKEELNDEKINVQDGTQIMEPIGLDSNGAEAQIDQ 835
QY 381 HCEPK-EGVTKIKASSLSTVKIABQPSYFPDPPPTFIFSPA-----NRRGR 428
Db 836 NITNETEILVAKPVSLDVKSVEQ-----MQPK--LESFVSSETSCTVDEKIEEK 887
QY 429 PPKRIHISQEDNVANKOTL-ASYRSKATKRDKLLKQEMKSLAFKAKLKR-----EKAD 483
Db 888 PEEVTLYQEGVDGSGYLETKEETVSPESIELEEQPBERSVIDTLPQKLTLESPSE 947
QY 484 ALEAKKEKEDK-EKKEE--ELAKVIEER---LKKKEKERLKVREKEREKREKREK 536
Db 948 VLEESSKTVEKTEETDSIELGIAQERSVTDLTPLQESSQPNQEQETK----- 1000
QY 537 KYVEYLKQWSPKPRDEMCDDLKELPEPTPVKTRLPPEIFGDALMLVLEFLNAFGE----- 590
Db 1001 ----LEKHEPTNEEVKSDEVIELSASPSK-----ELEGETVVEAENIENIKENEBEQA 1050
QY 591 -----LFDLQDEFPDGVTL--EVLREAL-----VGNDSSGPL 620
Db 1051 AEKIQKSLETVTQVESPSLLFSSEEQ--DHVTVAEEIVDEKAKEEVEPMLQIKNEDDA-- 1106
QY 621 CELLPFLTAIFQAIABEEBEVA-----KEQLTDADTKGCSLAKSLDLDSCT 666
Db 1107 ----TKIHETRVQARDIGPSLTEI CSINQNPQEEQVYKEACSEKEKEISTNSEN 1158
QY 667 L-SEILRLHILASGADVTSANAKRYQKRGFGFATDDACMELRSLNPSLVKKLSSTSYD 725
Db 1159 IVNETYALHVEAAEETATNG-----ESLDDVETTKSVLLEVRKEEAEAMKTTDAEPLD 1214
QY 726 LTPGEKMKILHALCGKLLTLVSTRDFIED--YVDILRQAKQOFPRELKAEQHRKEREBA 783
Db 1215 AIEKEELE-----TVKTVVQDAKIVNNEETTAHESSELKGDNHQEKNAEPVE 1261
QY 784 RIRKKE-EKLK-----PQEQMKKEKQKLEDEQNSTADISIGE-----ERE 827
Db 1262 ATQNLDDAEQISREVTVDTREADITEIKVQEGGPTVETPTIQ-GEDEISETSLKE 1300
QY 828 DFDTSIESKDTEQKE--LDQDMFTEDDDPGSHKRRGRGRKGNGQGFKEFTRQEQINCVR 885
```



```
Db 1321 EVQDS---SKDTEHEHVLBERDI-----PQC-----ETLKAEAVDTSTV 1356
Qy 886 E-----LITAD-----BEEA-----LKHORKEKELLEKTSIAIACTNIPPLGRDRMY 929
Db 1357 EEAAILKLELTIWISPEANGHSETSLDLKVKQKEAEIVKTV-----1398
Qy 930 RRYWIPSPISPL--FIEDYSGLTEDMLPRSSQNNVQSDPQVSTKGTGPLMSESTS 987
Db 1399 ----IFSNEVGTSDAQAEFGHTE-----PCSEIKDBSQSGSEESVEVSKETVQGESSE 1450
Qy 988 NTD-----QGRDRHSVOLPKP-----VHKPNWCYFSSCEQLDOLLEALNSGHESALK 1037
Db 1451 EKDVNMLDVQSGESSEKYQNEPDISLVSKT-----ENGDKFEIPSPVVEGAGLDETTHN 1504
Qy 1038 ETLLOEKSRICLAQLARFSEBEK-----HFSKDPQDSKPTYSRGRSSNAYDPPSQWCAEK 1091
Db 1505 QTLLOVESVKSLOTPTPSEETSKIDEKIDPK-EEVTLHQEGREESYG-----1555
Qy 1092 QLELRDLPLLDIEDRIYOGTLGAIKVTRDHRWIRSALESGRYELLSEENKENGIIKTVNE 1151
Db 1556 -----LTKDB-----AVSVLES-----RELGEOPQOEELCLANEQENETKLOEE 1595
Qy 1152 DV-----EEMEIDQTKVIIVKDRLLGKTETPTSTVSTNASTPOSVSVVHYLAKALFO 1204
Db 1596 QVDKHEPTKEEVNDQSPV-----EISNEVIQVSSASLSEGEPEYETVWEAKIGBEQ 1649
Qy 1205 IQOGIERF-----LKAPLDASDGRSYKTVLDWRRESLLSSAS-----LS 1245
Db 1650 VADKIQKSFETGEIWEAHSSLPSSSEKEHEVTSKTDDEKVDKDAEPTGDMRERGLDIAE 1709
Qy 1246 QVFLHSLTLDRSVIWSKSIILNARCKICRKGDAENNVLDGDCDRGHHTYCV-----RPK 1299
Db 1710 TTHLSLPSVD-----QKEDVDEI-----HIPSVALPLDEQEK 1741
Qy 1300 LKTVPEGWFCPCPKQRCRLSRFRQPSLESDEDDVDSMGGEDEVDGDEEGEQSEEE 1359
Db 1742 VTSEKGETKSSEAEEDDKPDEHVDSTSPMLSEKNDNETQTSKTSSEVCMQOEESGTLEV 1801
Qy 1360 EYEVQDEDDQSEEEVSLPKRGRQVRLPVKTR---GKLSSFSRSGQQQEPGRYPSS 1416
Db 1802 KPPESEKDKSQEISETIEEIBATSDQTLPIETSHDNTLSSELVSEQDDQSPKKV-BEI 1860
Qy 1417 QOSTPKTT--VSSKTRSLRKINSAPPTETKSLRIASRSTRSHSGPLOA-----1463
Db 1861 HEEEPKEAHDVEATSERNL-PVETSDADNTLSSQLVS-ETKEEH-KLQAGEITLPTIIPR 1917
Qy 1464 ----DVFVLLSPRRKRRGRKSANTTPENSFPNFRVIATKSSQSRSVNIATSKLSQE 1519
Db 1918 ESSDEALVSMLASREDDKVALQEDNCADD-----VRETNDIQEERSISVETEEVSQE 1969
Qy 1520 SESKR-----RCRKQSPSPSVTLGRSS-----GROGVHLSAFELQVYV-----ELV 1564
Db 1970 TPKKEHEBIDRAHVETP-TAPIILLENDSETLIAEAKKNGEENETERTVALDHEEBFV 2028
Qy 1565 RHD 1567
Db 2029 NHE 2031
```

RESULT 49

```
Q81J44
ID Q81J44 PRELIMINARY; PRT; 1596 AA.
AC Q81J44;
DT 01-MAR-2003 (TRENBLrel. 23, Created)
DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE Liver stage antigen, putative.
GN ORFNames=PF10_0356;
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
```

```
RP SEQUENCE FROM N.A.
RX MEDLINE=22255705; PubMed=12368864; DOI=10.1038/nature01097;
RA Gardner M.J., Hall N., Fung E., White O., Berrihan M., Hyman R.W.,
RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA Chan M.S., Nene V., Shallow S.J., Suh B., Peterson J., Angiuoli S.,
RA Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA Martin D.M., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA Fraser C.M., Barrell B.
RT "Genome sequence of the human malaria parasite Plasmodium
RT falciparum."
RL Nature 419:498-511 (2002).
DR EMBL; AE014834; AANJ5553.1; -.
SQ SEQUENCE 1596 AA; 191691 MW; C0B3A2CC0F5D60AC CRC64;
```

Query Match 3.6%; Score 313.5; DB 2; Length 1596;
Best Local Similarity 19.2%; Pred. No. 8.5e-05;
Matches 270; Conservative 263; Mismatches 567; Indels 305; Gaps 57;

```
Qy 263 KDRYFVEETVEIRNNGARLQCTILEVLPPSHONGFANGHVNSVDGETI--IISDSDSE 320
Db 80 KDK---ELTMSNVKVSQTNFKSLRNLVGS-ENIFLKNKLNKKGKLEHIIHNDKDK 135
Qy 321 TQSCSFQNGKKDAIDPLLFYKVQPTKKELHESAIVK---ATQISRKKHLFSRDKKL 376
Db 136 KYVGQDENRQEDLEQERLAKKLGQSQSDLEQERLAKKLGQSQSDLEQERLAKKLG 193
Qy 377 FLKHCEPQEGVIKASSLSTYKTAEQDFSYFFPDPTTFISPANRRRGPPRIHIS 436
Db 194 -----QEQQSDLEQDRLAKELQEQ-----SDLEQER-RAKELQEQ 230
Qy 437 QED---NVANKTTLASYRSKATKE-----DKLLK---QEEMKSLAFE--- 473
Db 231 QSDLERTKASTETLHEQQSDLEQERLAKKLGQSQSDLEQERLAKKLGQSQSDLEQERL 290
Qy 474 -KAKLKREKADALBAKKKEKED-----KEKREELKVIVEERLKKKEEKE 518
Db 291 AKELQEQSQSD-LEQERAKKLGQSQSDLEQDRLAKKLGQSQSDLEQERAKKLGQEQ 349
Qy 519 RLKVEREK-EREKURE-----EKRYKVEYLKQWSPKPREMECDLKLPEPTPVKTRL 570
Db 350 QSDLEQERLAKKLGQSQSDLEQERAKKLGQSQSDLEQERAKKLGQSQSDLEQERL 409
Qy 571 PPEIFGDALMVLEFLNAGFELFDLQDFPDGVTLEVLVEALVGNDSGELCELLFFFLTA 630
Db 410 AKELQEQSQSDLEQERRAKE--KLQEQSQSDLEQERLAKKLGQSQSD----- 454
Qy 631 IFQAIABEEBEEVAKELQ-----TDADTKGCSLKSLLDSDCTLSILRL--HILASGADVTS 684
Db 455 -----LEQERLAKKLGQSQSDLEQERAKKLGQSQSDLEQERAKKLGQSQSDLEQ 508
Qy 685 AN-AKRYVQKGGFDATDDACMEURLSNPSLVKLSSTSVVDLTLPGERKKIHLALCGKLL 743
Db 509 ERLAKELQEQ-----QSDLEQERLAKKLGQSQSDLEQERRAK-----EKLQ 551
Qy 744 TLVS---TRDFTIEDYVDILROAKQEPRE---LKAQEHKEK-----EAAAIRKKEEKLK 794
Db 552 EQQSDLEQERLAKKDLRLAKKLGQSQSDLEQERRAKKLGQSQSDLEQERRAKKLGQ 611
Qy 795 EQEQMKKEK---QEKLKE-----DEQRNSTADISIGEEEREDFTSIESK-----DT 838
Db 612 EQQSDLEQERLAKKLGQSQSDLEQERRAKKLGQSQSDLEQERRAKKLGQSQSDLE 668
Qy 839 EQKELQDMFTTEDDDPGSHKRGRRGKRGQNGKFEFTR-----QEINCVTVRELLTAD- 891
Db 669 EQERLAKKLGQSQSDLEQERRAKKLGQSQSDLEQERLAKKLGQSQSDLEQERLAKK 728
Qy 892 -EEEAALQEHORKEKELLEKIOSAIACTNIPLGDRMYRYWIYFPIPLFIEDYSGL 950
Db 729 LQEQSQSDLEQERLAKKLGQSQ-----DLQERLAKK-----LQEQSQSDLE 770
```



```
Db      1292 LLEDK-KOEKKYKEEIDOLGKENEDIITKONKELNRL-----EDYSAKIDAKOEELKLA 1345
Qy      838 -----TBQKELDQDMFTEDDEDPGSHKRRGRGKGQNGKFEFTQEQINCVCVTRRELTAD 891
Db      1346 NDVAVATKKKMKLEBKIKLEDT-----QHIFK-----DSENSLKSEL----- 1384
Qy      892 EEBALKQBHQKEKELLKISAIACINIFPLGRDRMYRRYWIPIFGLFIEDYSLT 951
Db      1385 EKTALMNNELSRDNDNIITKLNELORTN-----DKL-----IENKR--T 1422
Qy      952 EDMLLPRESSFONNVQSDPQVSTKTG-----EPLMSESTNIDQOPRDHSVQLPKPVHK 1006
Db      1423 EEKL-----RSEVAKUKDELKTSDTPEKRLMNNEDSSTI-----IKEYSEKI----- 1466
Qy      1007 PNRWCFYSCQOLDIIBALNSRGHRESALKETLLQEKSRIC-AQLARFSEBEKFHFSK- 1064
Db      1467 -----SSLEEKVETIKSEYDKEI-----NILEDKKEVLESELSKQKEIIDYNQKI 1512
Qy      1065 POPDSKPTYSRGRSSNAYDPQMCAEKOLELRDRFLL-----DIEDRIYQOTLGAI 1116
Db      1513 KEQETKAT-----EKEKEIQVAKNALNAEKKKKOIENDL-RTTIATV 1554
Qy      1117 KYTDRHIWR-SALES-----GRYELLSEE-NKENGIIKTWNEDVEEME-----I 1158
Db      1555 EKENTTLKRENOLKSESDIKHQNNTHLLOEELSQKELADKKHDBEIRKLEKENSXMDRI 1614
Qy      1159 D--EOTKVIVKIRLIGITETPTSTVSTNASPTQSVSSVVHYLAMALFOEQIERRFLKA 1216
Db      1615 DKLEKQKADTNEKIANIEKENSLLSERKTLVEKEN-----FQ-----DEITNLKS 1661
Qy      1217 PLDASDS-GRSYKTVLDNRWRESLLASSLSQVFLHLSTLDRSVIWSKSIILNARCKICRKK 1275
Db      1662 SLEKNDLSLSSHDELKDFNE-----LETELKRNLTEN-----K 1696
Qy      1276 GDAENMVLCDG-CDEGHHITYCVRPKLTVPEGDMWPCPCRCRQLRSPRPSLESDE 1334
Db      1697 LESENKQLSDKVIHEEKVSMVEKELSTA-----OKTLKERE 1733
Qy      1335 DVEDSMGGEDDVGDEEGOSEEEYE---VEQDEDDQBEVEVS-----LPKRG 1382
Db      1734 DVINKLDSNNELNKTIDKHGATEKHVEESITKKDSDIAQLKKIKDIEDKLSNILEKA 1793
Qy      1383 RPQVRLP--VKTRGKLSSPSSRQGOQEPGRYPSSQOSTPKTTVSSK--TGRSLRKINS 1438
Db      1794 KAAMLMTQLEKDKTDLKNSESELKQLESHYRSKYSSLESKLKSTEEAKKHVEESREQHQ 1853
Qy      1439 APPTETKSLRIASRTRSHGFLQA-DVVELLSPRRKRGKRSANNTPENSPNPNRV 1497
Db      1854 SMSLDLKATDKLKSAAETISISEMDAIKKQVELLT--KENVDLKSXNKADNS----- 1903
Qy      1498 IATKSSEQSRSVNIASLSLOESSESKRCRKRQSPSPVTLGRRSSRGQGVHLSAFE 1557
Db      1904 -----AKLSELDELKENEKLO-----LKNDRS-----ELDDL 1934
Qy      1558 QLVVELVRHDDSWPFLKLVSKIQVDDYDIITKPIALNIIREKVNKCEYKLASEFIDD 1615
Db      1935 LLVTELDEKNSKY-----REKLSLGVELSSDEEDD 1965
```

Search completed: March 9, 2005, 14:18:38
Job time : 237.583 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 9, 2005, 13:38:21 ; Search time 32.965 Seconds
(without alignment)
4885.990 Million cell updates/sec

Title: US-10-702-148-1
Perfect score: 8631
Sequence: 1 MEDASERGRVAPLNNVL.....VTPSNVDQVSTPPAAKKSRI 1674

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 65 summaries

Database : PIR 79: *
1: pir1: *
2: pir2: *
3: pir3: *
4: pir4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4168	48.3	811	2 T08738	hypothetical prote
2	945.5	11.0	1479	2 T17401	transcriptional regu
3	530	6.1	1430	2 T34516	hypothetical prote
4	418.5	4.8	449	2 T12495	hypothetical prote
5	397	4.6	1378	2 G88637	protein F53H1.4 [i
6	336	3.9	5327	2 T13564	microtubule-associ
7	319	3.7	3488	2 T34418	hypothetical prote
8	311	3.6	1549	1 A40691	trichohyalin - she
9	310	3.6	1898	1 A45973	trichohyalin - hum
10	309.5	3.6	1819	2 A1928	cag island protein
11	309	3.6	1526	2 A45605	mature-parasite-in
12	303.5	3.5	1130	2 T34081	hypothetical prote
13	303	3.5	678	2 A54514	glutamic acid-rich
14	298	3.5	2116	2 A26655	myosin heavy chain
15	297	3.4	2101	2 A42184	nuclear mitotic ap
16	295.5	3.4	1875	2 S38173	myosin-like protei
17	295	3.4	1909	2 A45592	liver stage antiage
18	294	3.4	1264	2 S64146	probable membrane
19	293.5	3.4	2663	1 S28261	centromere protein
20	289.5	3.4	2139	2 T18295	myosin heavy chain
21	286	3.3	1790	2 S67593	transport protein
22	285.5	3.3	1871	2 D96796	probable heat choc
23	285.5	3.3	1927	2 G64585	cag pathogenicity
24	283.5	3.3	2364	2 A65577	microtubule-associ
25	280	3.2	1269	2 F84730	probable myosin he
26	280	3.2	1407	1 S28589	trichohyalin - rab
27	280	3.2	1549	2 T21809	hypothetical prote
28	279.5	3.2	1961	1 A61231	myosin heavy chain
29	279	3.2	1827	2 T16270	hypothetical prote

ALIGNMENTS

RESULT 1

T08738

hypothetical protein DKFZp586E0518.1 - human (fragment)
C:Species: Homo sapiens (man)
C>Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 09-Jul-2004
C:Accession: T08738
R:Wambutt, R.; Heubner, D.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, March 1999

A:Reference number: Z16471
A:Accession: T08738
A:Molecule type: mRNA
A:Residues: 1-811 <WAM>
A:Cross-references: UNIPROT:Q9NRL2; EMBL:AL050089
A:Experimental source: adult uterus; clone DKFZp586E0518
C:Genetics:
A:Note: DKFZp586E0518.1
F:709-764/Domain: bromodomain homology <BRO>

Query Match 48.3%; Score 4168; DB 2; Length 811;
Best Local Similarity 99.4%; Pred. No. 1.4e-173;
Matches 806; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY	864	GKRGNGKPTROEQINCVTRRELLTADDEERALKQEHORKEKELLEKIQSAIACTNIPL	923
DB	1	GKRGNGKPTROEQINCVTRRELLTADDEERALKQEHORKEKELLEKIQSAIACTNIPL	60
QY	924	GRDRMYRYWIFPSIPGLFIEDYSGLTEDMLPRPSSFQNNVQSDPQVSTKTGEPLMS	983
DB	61	GRDRMYRYWIFPSIPGLFIEDYSGLTEDMLPRPSSFQNNVQSDPQVSTKTGEPLMS	120
QY	984	ESTSNIDGPRDHSVQLPKPVHKPNRWCFYSSCSOLDOLIIALNSGRHRESALKETLQOE	1043

Db 121 ESTSNIDQGRDHSVQLPKPVKPNRWCFSYSCQDQDQIIEALNSRGHRESALKETLQ 180
Qy 1044 KSRICAQALARESEKPFHSDPOPSKPTYSRGSSNAYDPSQCAKQELRLRDLFD 1103
Db 181 KSRICAQALARESEKPFHSDPOPSKPTYSRGSSNAYDPSQCAKQELRLRDLFD 240
Qy 1104 IEDRIYQGTGALGVTDRIHRSALSGRYELLSEENKENGIIKTVDNEDVEEMEIDEQTK 1163
Db 241 IEDRIYQGTGALGVTDRIHRSALSGRYELLSEENKENGIIKTVDNEDVEEMEIDEQTK 300
Qy 1164 VIVKRLIGIKTEPTSTVSTNASTPQSVSSVHYLAMALFOIEGIERRFLKAPLDASDS 1223
Db 301 VIVKRLIGIKTEPTSTVSTNASTPQSVSSVHYLAMALFOIEGIERRFLKAPLDASDS 360
Qy 1224 GRSYKTVLDRWRESLSSASISOVFLHSLTLDRSVWNSKSTILNARCKICRKKGAENMYL 1293
Db 361 GRSYKTVLDRWRESLSSASISOVFLHSLTLDRSVWNSKSTILNARCKICRKKGAENMYL 420
Qy 1284 CDGCDRGHHTYCVRPKLTVPEDGWFCPCRPKQRCRLSPQRPSLSDSDVDSMGGE 1343
Db 421 CDGCDRGHHTYCVRPKLTVPEDGWFCPCRPKQRCRLSPQRPSLSDSDVDSMGGE 480
Qy 1344 DDEVDGDEEQSEEEYEVEQDEDDSQEEVEVSLPKRGRPQVRLPVKTRGKLSFSSR 1403
Db 481 DDEVDGDEEQSEEEYEVEQDEDDSQEEVEVSLPKRGRPQVRLPVKTRGKLSFSSR 540
Qy 1404 GQOQEPGRYPSRSQOSTPKTVSSKTGRSLKINSAPPTETKSLRIASRSTRSHGPLEA 1463
Db 541 GQOQEPGRYPSRSQOSTPKTVSSKTGRSLKINSAPPTETKSLRIASRSTRSHGPLEA 600
Qy 1464 DVFVELLSPRKRRGRKSNANTPNSPNFPRVITATKSSQSRSVNIASKLSQESSEK 1523
Db 601 DVFVELLSPRKRRGRKSNANTPNSPNFPRVITATKSSQSRSVNIASKLSQESSEK 660
Qy 1524 RRCRKRQSPSPVTLGRSSRGQGVHLSAFEQVVELVRLHDDSWPFLKLSKIQVPD 1583
Db 661 RRCRKRQSPSPVTLGRSSRGQGVHLSAFEQVVELVRLHDDSWPFLKLSKIQVPD 720
Qy 1584 YYDIKKPIALNIREKNKCEYKLASEFIDDIHLMFNSCFEYNPRNTSEAKAGTRLOAF 1643
Db 721 YYDIKKPIALNIREKNKCEYKLASEFIDDIHLMFNSCFEYNPRNTSEAKAGTRLOAF 780
Qy 1644 FHIOAKLGLHVTSPNVDQVSTPPPAKKSRI 1674
Db 781 FHIOAKLGLHVTSPNVDQVSTPPPAKKSRI 811

RESULT 2
Ti17401
transcription regulator WBCR9 - mouse
N:Alternate names: Williams-Beuren syndrome deletion transcript 9 homolog
C:Species: Mus musculus (house mouse)
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T17401
R:Peoples, R.J.; Cisco, M.J.; Kaplan, P.; Francke, U.
Cytogenet. Cell Genet. 82, 238-246, 1998
A:Title: Identification of the WBCR9 gene, encoding a novel transcriptional regulator,
A:Reference number: 218735; MUID:99077764; PMID:9858827
A:Accession: T17401
A:Status: preliminary;
A:Molecule type: mRNA
A:Residues: 1-1479 <PRO>
A:Cross-references: UNIPROT:Q92277; EMBL:AF084480; NID:g4165088; PID:g4165089; PIDN:AADO
C:Genetics:
A:Gene: Wbcr9
A:Map position: 5
F:1360-1415/Domain: bromodomain homology <BRO>

Query Match 11.0%; Score 945.5; DB 2; Length 1479;
Best Local Similarity 22.9%; Pred. No. 1.6e-33;
Matches 393; Conservative 256; Mismatches 529; Indels 541; Gaps 62;
Qy 152 PLLHKKPVRQKPPADLPDEVEFYCKVTNEIFRHYDDFFERTILCNLSLVSCAVTGRGP 211

Db 3 PLLHKKPVRQKPPADLPDEVEFYCKVTNEIFRHYDDFFERTILCNLSLVSCAVTGRGP 59
Qy 212 LTYQEALESEKARQNL-QSFP---EPLIIPVLYLTSLTHSRHLHEICDDIFAYVQORYF 267
Db 60 LTHKEAEESEQAEALLKEBFPNWEKLVLEMH---HNTASLEKLVDSAWLIMTKYA 115
Qy 268 VEETVEVIRNNGARLOCTILEVLPSPHQNGFANGHANSVDGETI-----IISDSDSE 320
Db 116 VGEBCDFEVGKEMLVKVIKIH---LEKVDEAEVKKSDGACDPSDDKE 164
Qy 321 TQSCSFQNGKKDAI-----DP-----LLFKYKV 344
Db 165 NSSQMAQDLQKQETVVKDEGRRESINDRARRSPKLPSTLKKGERKWAPKPLPHKYDV 224
Qy 345 QPTKELHESAI---VKATQISRRKHLFSRDKLKFLKQHC-----EPQGVK---IK 392
Db 225 ---KLQNEKIIISNPADSLIRTPPNKEILFYFIHNLALRAGTGENAPWVVEDELVK 280
Qy 393 ASSLSYKIAQODFSYFPDPPTFFISPNRR---GRP-----PKR----- 432
Db 281 KYSLPS-----KPSDFLLDPYKMTLNPSTKRRNTGSPDRKPSKKPKRDSLSPLNP 334
Qy 433 ---IHI-----SQEDNV-----ANKQTLASYR-----SK 453
Db 335 KLWCHVHLEKSLNGPPLKVKNSKNSKSPBEHLEGVKIMSPNNKJLHSHFIPKGPAAKK 394
Qy 454 ATKERDKLLK-----QEBMSLA----- 471
Db 395 PGHSDPLKAKGRGKILANGQKSTGNSKSPSKVCTPKTKMKQMTLLDMAKGTQKWT 454
Qy 472 -----FEKAKLREKADALE-----AKKEKEDKEK 497
Db 455 PRSGGVPRSSGKPHKHLPPAALHLIAYYKENKDKEDKKSALSCVISKTAALLNEDRAR 514
Qy 498 KREELKKIVBEE-----RLKKKEKERLKVREKEREKLR---EKRTVVEYLKQ 544
Db 515 LPEELRALVQKRYELHEHKKRWASMSBEORKEYLKKRQELKERLREKAKERRERMLER 574
Qy 545 WSKPR---EDMECDLKBELPPTPVKT---RUPPEIFGDALMVLEFLNAPFGLDQDEPDP 600
Db 575 LEKQKRFDEQBLGG-RNLPAFLVDTPEGLPNTLFGDVALVVEFLSCYSGLLLPDAQYP- 632
Qy 601 GVTLEVEEALVGNDSGGLCELLFFELTAIFQAIABE---EEVAKQLTADATKGCGLK 658
Db 633 -ITAVSLMEAL-SADKGG-----FLYLNRLVLLQTLQDEIAB-----DYGELGMLKS 680
Qy 659 SLDLDSCTLSLRLHLILASGADVTSANAKYRQKGGFDATDDACMBELRLSNPSLVK 718
Db 681 EIPLTLSHVSSELVRLCRRCDVQEDSGSETD-----DNKDTSTPPEDNEVQDEFLEKL 733
Qy 719 SSTSVYDLTPEGKWKILHALCGKLLTLVSTRDFTEDYVDILRQAKQEFRELKASQHRKER 778
Db 734 ETSEFFELTSEKLRILTLCHRLIMTYSVD-----HMETR 770
Qy 779 EEAARIRKREELKEQEQMKKEQKLEKLEDEORNSTADISIGEEREDFDTSIESKDT 838
Db 771 QQVSAELWKEARLAVLKEENDKRAEKQKREMEARNK-----ENGKEE-----NVLGKVD 820
Qy 839 EQKE---LDQDMFTEDEDDPGSHKRRGRKRGKQNGKFETRQEQINCQVTRRELLTADEEA 895
Db 821 RKKEIVKIEQOQVEADDMISAVSRLLSNQAKREIKERETKVLERA-----EER 876
Qy 896 LKQBHQKEKELKEIKISAICTNIFPLGRDRMYRYWIFPS-IPGUFIE----- 944
Db 877 MRKHAATAAFAFGIAKAVLIRRTPIGTDRNHNRYLFSNEVPGFLIEKGVHNSIDY 936
Qy 945 -----EDYSGLTEDMLLPRPSSQNNVQSQDPQVSTKTGEPMLSESTNSIDQGRHSHV 998
Db 937 RFKHKRKHDSNLPDDCYCPR-----RKKANLGNKASVNAHHPALEAV 979
Qy 999 QLPKPVHKPNRWCFSYSCQDQDQIIEALNSRGHRESALKETLQEKSR---CAQLARFSE 1056


```
Db 980 ETTVPKQGNLWFLCDSQKELDELLSCLHLPQGIRESQFKERLEKRYOBITHCNVYAR --- 1036
Qy 1057 EKHFSDKPDQPSKPTYSRGRSSNAYDPSQMAEKAQLELRDRFLDIEDRIYQOTLGAI 1116
Db 1037 -----KPNLGLK-----SCDGNQ-----ELLNFLRSLDIEVATRLQKGLGYM 1074
Qy 1117 KYTDRHWRSALESGRYELLSEENKENGIIKTVDNEVEEMEIDEQTKVIVK--DRLLGIK 1174
Db 1075 EGT-----SEPEA--RVISLEKL-----KDFGECVIALQASVIKFLQGFMAPK 1116
Qy 1175 TETPSTVSTNASTPOSVSVVHYLAMALFQIEQGIERRFLKAPLDASDGRSYKTVLDRW 1234
Db 1117 OKKRKLOESDTSKSEVDE-----EKWVEAKVAS-----ALEKW 1152
Qy 1235 RESLSSASLSQVFLHSLTDRSVTSWKSILNARCKI CRKKGDAENMVLCDGCDRHGHTY 1294
Db 1153 KTAIREAQTFSRMHVLLGMLDACIKWMSAENARCKVCRKKGEDDKLILCDECNKAFHLF 1212
Qy 1295 CVRPKLKTVPEGDWCPCEPRKQRCRRISFRORPSLESDDEDVEDSMGDEDDVDGDEEG 1354
Db 1213 CLRPALYEPDGEWQCPACQPTARNSRGRNYTEESTSEGSEGESEEEE--EEEE 1269
Qy 1355 QSEBEYEVEQDEDDSQBEEVSLPKRGRPOVRL-PVKT-RGKLSFSFSSRQOQEPGRY 1412
Db 1270 EEEEDYEV-----AGURLRPRKIIRKQSVIPAAR-----PGRP 1304
Qy 1413 PRRSQOSTPKTTVSSKTCRSLRKINSAPPTETKSLRIASRSTRHSHGFLQADVVELLSP 1472
Db 1305 PGKK-----SH-----P 1311
Qy 1473 RKRRGRKSANNTPNSPNFNRVIAKYSSRQSRVSNIAKLSQESKRKRCKRQSP 1532
Db 1312 ARRSR-----PKDDPEVDOL-VLQTK-----RISRRQS- 1338
Qy 1533 BPSVTLGRSSRGQGVHLSAFQVLVVELVRHDSWPFKLKLSKIQVPDYDIIKKPI 1592
Db 1339 -----LELQKEDILHUKLVYRFSWPFPRVTRDEADYDVIEHPM 1380
Qy 1593 ALNIIRKVNKCEYKLASEFIDDIELMFNSCFEYNPRNT 1631
Db 1381 DFQTIQKSCGNYRSVQEFELTDMKQVFANAELYNCRGS 1419

RESULT 3
T34516
hypothetical protein ZK783.4 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 09-Jul-2004
C:Accession: T34516
R:Favell, A.; Vaudin, M.
A:Description: The sequence of C. elegans cosmid ZK783.
A:Reference number: Z21536
A:Accession: T34516
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1430 <FAV>
A:Cross-references: UNIPROT:023590; EMBL:U13646; PIDN:RAC24421.1; GSPDB:GN00021; CESP:ZK
A:Experimental source: strain Bristol N2; clone ZK783
C:Genetics:
A:Gene: CESP:ZK783.4
A:Map position: 3
A:Introns: 248/3; 373/3; 547/1; 593/2; 905/2; 1042/3; 1116/3; 1317/2; 1376/1
F:1284-1339/Domain: bromodomain homology <BRO>

Query Match 6.1%; Score 530; DB 2; Length 1430;
Best Local Similarity 20.1%; Pred. No. 1.6e-15;
Matches 304; Conservative 194; Mismatches 502; Indels 516; Gaps 54;
Qy 296 NGFANGHNSVDGETIILISDSDSETQSCSFQNGKKDAIDPLLFKYKVPQTKKELHESA 355
Db 214 NNAASNMMNNVWQLVAQMOKQOQOQKQDQTO--KKADQA-----KKAKELAQOQKQEQD 266
```

```
Qy 356 IVKATQISRRKHLFSRDKLKLFLKQHCPEQGVIKIKASSLSYKIA----- 402
Db 267 VKNQOQEEILFLMAHQHNLN-HOKKHEKQADAAALAAKVAALAAHRAALLESDESPEBKKTN 325
Qy 403 -----EODFSVFPPDDPTFTFFSPANRRRRGPPPKRIHIS 436
Db 336 EAMLELPLOLQWRROTCTVRSASAGVKGDVSYFAPCGKKLSTYSEVRYLTG--NSIHYI 383
Qy 437 QEDNVANKOTILASYRSKATKERDKLLKQEMKSLAFKAKLKRE----- 480
Db 384 TRDNFLNTKLVIQEFIVPKTEADETQOEREFAMFTEDDINKELTRLNLVLFVPKIQAS 443
Qy 481 -----KADALAKKKEKEDKKEKREELKKIIVEERLKK--EERERLAKVEREKREKL 532
Db 444 TSNGVHEDDIKMSKIEEPDEPLDPSELNDEFTIELVHSQIMNSNGVDECKI-RERADDLL 502
Qy 533 EEKRYVEYLQWKSPPREDMCCDDLKELPEPTPVKTR-LPPEIFGDALMVLFLNAFGBL 591
Db 503 -----VNINDVRHLPDFSRIGNQCLSSQGFADALMVHVEFVQNFHV 543
Qy 592 FDLQDEFPDGVTLV---LEBALVGNDSGLCELLFFFLTAIFQAI AE EEEBEVAKEQIT 648
Db 544 L-----GIDLEIAPKLESICAGLDGDANHAQET-LQLTRQLRLALEFPGMGNEK-- 592
Qy 649 DADTKGCSKSLDSDCTLSEILRHILIASGADVTSANAKYRQYKRGQGFDDDDACMEUR 708
Db 593 ---RFGGGGEMGLDRENFSEVMRLFLIDKG-----KRGE-----E 625
Qy 709 LSNPSLVKKGSLSTSVYDLTPGCKMKILHALCGKLLTLVSTRDFIEDYDYLILQAKQEPRE 768
Db 626 LSQP-----LUTCNPLSISPEQKASILAFLODE--LVCSRNVVTE--IDKNLDEISR 673
Qy 769 LKAEQHRKE---REAAARIRKREK---LKEQKQMKKEQKELKEDEQRNST----- 816
Db 674 LKGEKMWEGKARALRSARKKNDKVVVKEEQNHESDSEPTRPDPKPKATVAPTV 733
Qy 817 ---ADISIGEEREDFTSIESKQTEKELQDMFTEDEDDPGSHKRRGRKRGQNGFK 873
Db 734 VSVSPVSAAOQQRKF-----TPG-----LQO----- 755
Qy 874 FTROEQINCVTRRELLTADDEEALKOE-----HORKEKELLEKIQSIACTNIFPLGR 925
Db 756 -----C-----EVLTEQESMSIQOMDSLIGDLHQ-EAQNINQKHHTGLKIRFPFGT 803
Qy 926 DMVRRYWIPIPSIGLFI EEDYSGLTEDMLPRSSFNQNVQSOPQVSTTKTGELPMSES 985
Db 804 DRFHRYWMLAHTDKVILIES-----LATSVNNPACNANEVASKDP-----PTEQRVPCA 854
Qy 986 TSNIID-----QGRDRHSVOLPKPVHKPNR-----WCFYSSCEQLDOLIEA 1025
Db 855 CETIDLQVIACVEDLVDDVLLRAKADKTKRKYRRIENHMKRGWMTQNRDCVESLRSC 914
Qy 1026 LNSRGHRESALKETILQ-----EKSRI CAQLAR-----FSEERKF 1059
Db 915 MLRSGIRERALHRLTLTKPWFNLKFGTITIEPVGEKSDL--ELVRQGWTRLNTAIDKL 972
Qy 1060 HFSKDPQSDQPTYSRGRSSNAYDPSQMAEKAQLELRDRFLDIEDRIYQOTLGAIKVT 1119
Db 973 QCHLKMSDVSRPL-----PSITPFTQKPI-----VVPTMALAQIV 1009
Qy 1120 DRHIWRSALSGRYELLSEENKENGIIKTVDNEVEEMEIDEQTKVIVKDRLLGIKTETPS 1179
Db 1010 -----KDDMAWKVIDEVDGQELDE---TIIRQKI----- 1036
Qy 1180 TVSTNASTPOSVSVVHYLAMALFQIEQGIERRFLKAPLDASDGRSYKTVLDWRRESLL 1239
Db 1037 -----IETADMDDQDTCOLFEDWKSYYVS 1058
Qy 1240 SSA-SLSQVFLHSLTDRSVTSWKSILNARCKI CRKKGDAENMVLCDGCDRHGHTYCVRP 1298
Db 1059 TEAQTTSQMLVALQTLQEGNIMWRSRREALCQIC-KSMGDGEMLVCDGCESGSCHNECFRP 1117
Qy 1299 KLKTVPEGDWFCPECRKQRCRRISFRORPSLESDDEDVEDSMGDEDDVDGDEEGQSEE 1358
```

```
Db 1118 RMTKYPEGDFC-----QRCR----- 1133
QY 1359 BEYEVEQEDDDSQEBEEVSLPKRGRPOVRLPVKTRGKLUSSFSRGGQQEQEPGRVPSRSQQ 1418
Db 1134 -----EE-----KSGRPMCWFCSRETGNLHQC-----QRCAYHVHQECQ 1168
QY 1419 STPKTTVSSKT---GRSLRKINSAPPTETKSLRIASRSTRSHGHPLOADVVELLSPRK 1475
Db 1169 DGPKEAINPETFCGHC-----QEMKQMR-----FVKRLIDRSE 1202
QY 1476 RRGKR-SANNTPENSPNPNFRV-----IATKSSEQRSV--NIASKL-----SQES 1520
Db 1203 SEERELEDNNAENGENTKNGHMGMNGAIALGVHNOQGVKGNLKRKLEVPISGGLPKN 1262
QY 1521 ESKRRCRKQPEPSVPTLGRSSRGQGVHELSAFQOLVVELVRHDDSWPFLKLVSKIQ 1580
Db 1263 MNKELC-----QMLDELVVQANALPFLBPVNPKL 1292
QY 1581 VPDYYDIITKPIALNIREKYNKCEYKLASEFIDDIELMFSCPEYNPRNTSEAKAGTRL 1640
Db 1293 VPGYKMIISKPMWDLTKIQKNEKLIYTPEDFAEDIELMFANCRQFNIDHSEIGRAGISL 1352
QY 1641 QAFFHIQAQKLGHLVT 1656
Db 1353 HKFTQKRWKQLKYNFT 1368
```

```
RESULT 4
T12495
hypothetical protein DKF2p434H07l.1 - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 23-Jul-1999 #sequence_revision 23-Jul-1999 #text_change 09-Jul-2004
C:Accession: T12495
R:Poustka, A.; Klein, M.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, June 1999
A:Reference number: Z17525
A:Accession: T12495
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-449 <POU>
A:Cross-references: UNIPROT:Q9UIF8; EMBL:AL080173
A:Experimental source: adult testis; clone DKF2p434H07l
C:Genetics:
A>Note: DKF2p434H07l.1
C:Superfamily: transcription factor GGN5; bromodomain homology
F:366-421/Domain: bromodomain homology <BRO>
```

```
Query Match 4.8%; Score 418.5; DB 2; Length 449;
Best Local Similarity 22.2%; Pred. No. 2.8e-11;
Matches 143; Conservative 93; Mismatches 181; Indels 227; Gaps 18;

QY 1017 EQDQLIBALNSRGHRESALKETL---LQEKSRICAQ-----LAFSBEKFHFSDKPQPD 1068
Db 6 EDLKALLKVLHLRGIREKALQIQKHLDYITQACLKNKDVAITELNENE---ENQVTRD 62

QY 1069 SKPTYSRGRSSNAYDPSCWCAEKQLELRDLRFLDIEDRIYQGTILGAIKVTDRHIWR-SA 1127
Db 63 IVENSVEEQAMWMDLSVLQOVEDLERRVASASLQVKG-----WMCPE 105

QY 1128 LESGRYELLSENKENGIIKTVNEDEVMEIDEQTKVIVKDRLLGIKTETPTVSTNAST 1187
Db 106 PASEREDLVVYEHK--STFKLCKEHGDEFTGEDESSAHALER----KSDNLDIAVT--- 156

QY 1188 PQSVSSVHYLAMALFOIEQGIERRFLK--APLDASDSGRSYKTVLDRWRSLSSASLS 1245
Db 157 -----RLADLERNIERRIEEDIAP-----GLRVWRRLSEARSAA 191

QY 1246 QVFLHLSTLDRSVIWSKSIILNARCKIKRKGDAENMWLDCGDRGHHTYCVRPKLKTUPE 1305
Db 192 QVALCQLOKSIAMWKSIMKYVCOICRKGDNNEILLCLDCGDCGCHTYCHRPKITTPD 251

QY 1306 GDWFCPECRPKQ-----RCRLSLFRQRPFSLESDVEDVDSMGDEDEVGDDEEGQSBE 1360
```

```
Db 252 GDWFCPACIAKASQTLKIKKLHVKGKKTNESK-----GKVVTLTGDTF----- 296
QY 1361 YEVEQDDEDDSQEBEEVSLPKRGRPOVRLPVKTRGKLUSSFSRGGQQEQEPGRVPSRSQQ 1420
Db 297 -----DBD----- 299
QY 1421 PKTTVSSKTGRSLRKINSAPPTETKSLRIASRSTRSHGHPLOADVVELLSPRKRGKRK 1480
Db 300 -----SASTSSSL-----KRNK 312
QY 1481 SANNTPENSPNPNFRVIATKSSSQSRSVNIASKLSLQESSESKRRCRKQPEPSVPTLG 1540
Db 313 D-----LKKRMEENTSI-----LSQESFSTVKPKRDSK----- 345
QY 1541 RRSRGQGVHELSAFQOLVVELVRHDDSWPFLKLVSKIQVPDYDIIKPIALNITREK 1600
Db 346 -----DLALCSMILTEMETHEDAWPFLPVNLKLVPGYKVKIKKPMDFSTIREK 394
QY 1601 VNKCEYKLASEFIDDIELMFSCPEYNPRNTSEAKAGTRLQAPF 1644
Db 395 LSSQYPNLETFAJDVRLVDFDNCETTFNEDDDSDIGRAGHNMKYF 438
```

```
RESULT 5
G88637
Protein F53H1.4 [imported] - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 09-Jul-2004
C:Accession: G88637
R:anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biology
A:Reference number: A75000; MUID:99069613; PMID:1981916
A>Note: see websites genome.wustl.edu/gsc/C.elegans/ and www.sanger.ac.uk/Projects/C.ele
A>Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
A:Accession: G88637
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1378 <STO>
A:Cross-references: UNIPROT:O45075; GB:chr_IV; PIDN:AAC02578.1; PID:g2854159; GSPDB:GN001
C:Genetics:
A:Gene: F53H1.4
A:Map position: 4
```

```
Query Match 4.6%; Score 397; DB 2; Length 1378;
Best Local Similarity 19.7%; Pred. No. 8.8e-10;
Matches 284; Conservative 224; Mismatches 524; Indels 412; Gaps 62;

QY 232 PEPLIIPVLVLTSLTHSRSLHEICDDIFAYVVKORYFVEETVEVIRNNGARLQCTILEVLP 291
Db 50 PKPLAIEVY---SKVHNSNLSE--DDLDEIED--LIDEKF----- 83

QY 292 PSHQNGFANGHVNSVDGETIIISDSDSETQSCSFONGKKKDAIDPLLFKYKQVPTKKEL 351
Db 84 -----PAGEKLMSSGKEYTVVSSEKRGGLTLYTMEDGK----- 117

QY 352 HESAIIVKATQISRKKHLFSDRDLKLFKQHCPOEGVIKKASSLSYTKIAEQDFSYFFP 411
Db 118 -----IGHRLRRKKGLSVEEIKKIAIEDAEAFVDEKQWKVREELLAENPIRE--IKKYAP 170

QY 412 DDPTFTFISPANRRGPPPKRIHISQEDNVANKOTLASYSRKATKRDKLKQEMKSLA 471
Db 171 -----IFS-ANRKSTPTKTAQLSVAADSDSDVOEVS-----MDAESGA 209

QY 472 FEKAKLKREK-----ADALEAKKKEKEDKKEELKKIVVEERLKKKBEKE 518
Db 210 ANKSAMKTPRCAPRASGGPVILSSRLQEKQEKDEKEKLE-KKQEQEKKQKKEE 268

QY 519 RLKVEREKREKLBEKKRVVEYLKQNSKPREDMCEDDLKELPEPTPVKTRL--PPEITGD 577
Db 269 KAKKLKEKE-EKLKEKEKAARKEKKNNGTMDKFLKKTGTGSPSSKNAPLFSFSKWGE 327
```

```
QY 578 ALMVLEFLNAGELFDLODEFDGVTVLEVEA-----LVGNDS-----EGPLCEL----- 623
Db 328 KRIANG-VKMDAMKRRD-----LELYNEACSWCKNLSGNORSTFPIKFSVQK 379
QY 624 -----LFFFLTAIFAIAEER-----EVAKE----- 645
Db 380 LVDKAKDRAHMGKWAQKAEKFAESEKRELYQKFEPIKAWFNEDIALDLDLVTCE 439
QY 646 -QLTADTKGCS---LKSIDLSCSTLSEILRLHILASGADVTSANAKYRYQKRGFDATD 701
Db 440 LELLDAKRVNCEDELLKCLEISQFFVS---MRKILLNENITA---EQLRDDLHCGFDG-- 492
QY 702 DACMELRLSNPLSVKXLSVYDITPGEKMKILHALCGKLTIVSTRDF-IEDYV--DI 758
Db 493 -----FKRSTYKMIANLLETA---LOEKEHEKAH--CNARLS-----EPPINEHITSEL 537
QY 759 LR-----CAQEFRELKAEQKEREAEAAARIKREKELKEQEQKMEKKEKLE 809
Db 538 IRAFFIGTSTFKRDGKRAARGAHDDEEDG-----DEEWDMSREKVVEIEKEP 590
QY 810 DEQRNSTADISIGEBEREDFTSISKD-----TEQKE----- 842
Db 591 NVENGAAENGIGDDEEENSESELQKORILALFADSCHYELPAGAQLVLCAMKEV 650
QY 843 --LDQDMFTEDDDPGSKRGGRGKQNGKFEFTREQ-----INCVTRELITADEEAL 896
Db 651 LPRIEFLWLRDANSEKLTETHEKKEANRINEMEQFHQLQDFPPIPEITESMTRT 709
QY 897 KOEHORKEK---EL-----LEKIOSAJA-----CTNIFPLGRDMYRR-YWIFPSI 938
Db 710 EQLRRREKLENLQTLKIELENREATAREVDDLERIFRVVHIGNDRHLRYKYWFAYSS 769
QY 939 PGLFIEEDYSGIATEDMLPRPSSFQNVQSDPQVSTKTGPELMSGTSNIDQGRDHSV 998
Db 770 DAAIWWQDF-GTT-----SYEKWYDCSEKFMVDSESDVENREYEDL 812
QY 999 QLPKVPKPNRWCFSYSSCEQLDQLEALNSRGHRESALKETLLQEKSRICHAQLARFSEK 1058
Db 813 PLTSS-QSSETWYKLDTEPAIRQLMTQKNGREKLLKYLKLRNMNDIISILR-KEKK 870
QY 1059 FHFSKPKQPDSK-----PHYSGRSSNAVDPSOMCAEQLELRDLDF 1100
Db 871 QKKDGEDEEASEDEASAAENGEMKETEONGTKFGPEPTAAAAAETQBAEKN- 929
QY 1101 LLDIEDRIYQGLGAIKVTDRHIMRSALSGRYELLSEBKENGIIKTWNEDVEEMIDE 1160
Db 930 -----ARFTGFGSLKRT-----MSELNDW-KQSGLSKIVDQVFEARMLE 971
QY 1161 QTKVIVKDRLLGIKTETSTVSTNA---STPOSVSVVHYLAMALFOIEQGIERRFLKAP 1217
Db 972 ANTL---DEMRLVTELVTISIPVECVIEKFPONVAIAKKFLNFTILLEKWKVNEF--- 1024
QY 1218 LDASDGR-----SVKT---VLDWRRESLLSSASL-----S 1245
Db 1025 ---SDFRAKCRIFNYLLFSHTKSKFLEKLNQTKMIGAKFGICAIFQKPDVRLLPKK 1081
QY 1246 OVFLH-----LSTLDRSVIWSKSLNARCKIKRKKGDANNVVL 1283
Db 1082 ECFSHLKMPRFCRRVQVEASNASCLHMLLAYFDARDQORTLPESLCQVCRKTGTGERK 1141
QY 1284 CDGCDRGHHTYCVRPKJKTIV---PEG---DWFCPECRPKQRCRLSRFRQPSLESDVE 1337
Db 1142 CKQCSTVFHYGCHRP TISRALFEGBGFKGWWCAKCTKEDRRRLQS-----BAKEDLR 1194
QY 1338 DSMGDEEDVDGDEGQSEBEEYEVEQDEDSQEBEVSUPLKRGPOVRPLVKTRGKLS 1397
Db 1195 -----QKEGGGDEEDHHGSGESDEDEEDIEVEETRGSAR---KANAAMR 1239
QY 1398 SFSFSGR-QQOEPGRYPRSQQSTPKTTVSSKTSRLKINSAPTEKSLRI---ASRS 1453
Db 1240 DVLEFEGVLRQTAPPPPRQR-----KTVVVEVLFLFNSIBRANRLYKTLIQIIPGQSR 1295
QY 1454 TRHS 1457
```

Db 1296 TRNA 1299

RESULT 6

T13564

microtubule-associated protein homolog - fruit fly (*Drosophila melanogaster*)
N:Alternate names: hypothetical protein EG:49E4.1

C:Species: *Drosophila melanogaster*

C>Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 09-Jul-2004

C:Accession: T13564

R:Spanos, L.; Papagiannakis, G.; Siden-Kiamos, I.; Louis, C.

submitted to the EMBL Data Library, April 1999

A:Description: Sequencing the distal X chromosome of *Drosophila melanogaster*.

A:Reference number: Z17689

A:Accession: T13564

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-5327 <SPA>

A:Cross-references: UNIPROT:O76891; EMBL:AL031128; PIDN:CAA20006.1

C:Genetics:

A:Cross-references: FlyBase:FBgn0025392

A:Introns: 24/2; 52/3; 104/3; 179/1; 232/1; 1669/3; 2566/1; 4798/3; 5272/1

A:Note: EG:49E4.1

C:Superfamily: *Drosophila* 576K microtubule-associated protein homolog

Query Match

Best Local Similarity 3.9%; Score 336; DB 2; Length 5327;

Matches 354; Conservative 252; Mismatches 679; Indels 522; Gaps 78;

QY 4 ASRSSR-GVAPLINNVLPGLSPISLPVSVT-GCKSHR--VANKKVEA----- 46

Db 3005 SQEASRPSEAESEKDAAPQSETSRPESVTESVKDGKSPVASKEASRPASVAENAKDSA 3064

QY 47 -RSEKLLPTALPPSEPK--VDQKLPRSS-----ERRSGGGTQPPASRAVAAGEA 94

Db 3065 DESKEQRPESLPQSKAGSIKDEKSLASKBAEKSKEESRESVAEQPLVSKVSRPAS 3124

QY 95 AARGAAGPERGSLGRRVSPRLCSGEGGVAVGVIAKG---RGRRGDGRRAPGGREM 151

Db 3125 VAESVKDEAEKS---KESRP--LMSKEASRP--SVAGSVKDEAEKSKEESRESVAES 3177

QY 152 PLLHRKPFVRQKPADLRPDDEVFYCKVTNEIFRHYDDFFERTILCNLSLVMSCAVTGRPG 211

Db 3178 PL-----PSKEASRPASVAESVK----- 3195

QY 212 LTYQEALESEKARQNQLSQSPPEPLIIPVLYLTSLTHRSRLHEICDDIFAYVKDVFVBET 271

Db 3196 ---DEADKSKESRESREGAEKSP-----ASKEASRPASVAESI-----KDE--AEKS 3238

QY 272 VEVIRNNGARLQCTILEVLP-PHQNGFANGHVNSVDGETIISDSDDSETQSCSFQNGK 330

Db 3239 KEESRRE-----SVAESKPLPSKEASRPSTVAESVKDE-----AEKSKEES 3279

QY 331 KKDAI---DPLLKYKQVPTK-KELHESAIVKATQISRRKHLFSRDKLKLFLKHCEPQE 386

Db 3280 SRDSVAESKPLASKEASRPASVAESVQDEAEKSKEESRESV--AEKSPLAYKEASRPAS 3337

QY 387 GVIKIKASSLSTYK-----IAQDFSYFFDDPPTTFIFSP-ANRRRRORP----- 429

Db 3338 VAESIKDEAEKSKEESRESVAEK-----SPLASKEASRPSTVAESVKDE 3382

QY 430 -PKRIHISQEDNVANKQTLASYSR-----SKATKERDKLLKQEMKSLAFKAKL-KRE 480

Db 3383 AEKSKEESSRSVAESKPLASKEASRPASVAESVQDEAEKSKEESRESRESVAESKPLASKE 3442

QY 481 KADALEAKKEKEDKEKKREELKK--IVEEERLKKKEK-----ERLKVREKEREKLR 532

Db 3443 ASRPASVAESVKDAAEKSKEESRESRESVAESKPLASKEASRPASVAESVKDAAEKSKEESR 3502

QY 533 EEKKRYEYVLKQWSEKPRDEMCDDLKELPEPTPVKTRLPPEIFGDALMVLFLNAGBLF 592

Db 3503 RESVAESKPLPSKEASRPSTVAESVKDAAEKSKEESR-----RESVAEKSSLA 3550

Db 885 RESQFQDLRP-----LQDEQEKREKREQWRQKRDQSFP 921
Qy 771 AEQ--HRKEREAARIKRKEE-----KLKEQKQMK--EKQKEL 807
Db 922 AEQLLRSQKTERDRKFRREEEQLLKGQREKIRYLEEDRKFRREEQQLRLREEQOL 981
Qy 808 KEDEQNSTADISIGEEEREDFTSIESKDTQEKELDMFTEDDDPGSHKRRGRKRG 867
Db 982 QERDRKFRBELSRQERD-----KFREEEQLO-----EREQLRQERDRKFR 1028
Qy 868 QNGKFETRQEQINCVTR-----ELLTADDEEALQKHQKREKE 1074
Db 1029 EQLLQE--REEQLRQERDRKFRREEEQLLQEREQLRQERDRKFRREEEQLLRLREEQ 1086
Qy 914 AIACINIFLGRDRMYRWYFPPGLFIEDYSLGTEDMLPPSPFQNNVQSDPOV 973
Db 1087 -----LRQERNRFR-----EQLL-REREQLRQEGEPOL 1117
Qy 974 STKTG-----EPLMSESTNIDQPRDHSVOLPKPVHKPNRWCFYSSCEQLDQLIEALN 1027
Db 1118 RQKDRKFHEEQLLQEREQLRQERDRKFRFEEAIIK-----EREQLRQER 1167
Qy 1028 SRGHRESALKETLLQEKSRIC-----AQARFSEKHFHSDK-----POPDSKPYTS 1074
Db 1168 DRKFRE--EEQLLQERBELRQEREPQLRQERDRKFRREEEQLLQEREKLRQEREPQLR 1224
Qy 1075 RGRSSNAYDPSQMAEKQLELRDLFLDIEDRIYQGTGALKVTDRIHWSALESRYE 1134
Db 1225 QERDRKFHEEQLLQEREQLRQER-----RDRKFRE-----EAQLLQEREE 1266
Qy 1135 LLSSENKENGIIKTNEQVVEEIMEIDQTKVIVKDRLLGIKTETPTSTVNSTAPOSVSSV 1194
Db 1267 QLRRQERD--KFREEQLLQEREQLRQERD----- 1298
Qy 1195 VHYLAMALFQIEQGIERRPLKAPLDASDSGRSYKTVLDRWRRESLSSASLSQVFLHLSTL 1254
Db 1299 -----KFREEQLLQER--EQLRQERDRKF-----REEEQLLKE--SEEQLRQER 1342
Qy 1255 DRSVTWSKSIILNARCKICRKGDAENNVLDGCDRGHTYCVRPKLKTVPEGDFCPCPR 1314
Db 1343 DRKFHEKHLRLREEEQLLQERQLEGV-----FSQEQLRRAEQEE--EQR 1387
Qy 1315 PKQRCRLSFRQPSLEDEVEDSMGGEDEVGDGEGQSEBEYVEQ-----DEDDS 1370
Db 1398 QRQDRKF--LEEQLLQEREEKRRVQEQDKFLEQEQLEHREQLHRRQLDQYR 1446
Qy 1371 QEEEVSLPKRGRPVQL---PVKTRGLSSFSF-----RQQQEPGRYPSRSQQSPK 1422
Db 1447 ABEQFARBEKRRRQBELRQERQRRQERERKFRREEEQLRQEQEQEKRRQERDVQQR 1506
Qy 1423 TVVSKTGR-----SLRKINSAP 1440
Db 1507 QVMEEDKGRQVLEAGKQFASAP 1530

RESULT 9
A45973
trichohyalin - human
C:Species: Homo sapiens (man)
C>Date: 03-May-1994 #sequence_revision 01-Mar-1996 #text_change 09-Jul-2004
C:Accession: A45973
R:Lee, S.C.; Kim, I.G.; Marekov, L.N.; O'Keefe, E.J.; Parry, D.A.D.; Steinert, P.M.
J. Biol. Chem. 268, 12164-12176, 1993
A:Title: The structure of human trichohyalin. Potential multiple roles as a functional E
ed (cross-linking) protein.
A:Reference number: A45973
A:Accession: A45973, MUID:93280194; PMID:7685034
A:Molecule type: DNA
A:Residues: 1-1898 <LSE>
A:Cross-references: UNIPROT:Q07283; GB:I09190; NID:9292835; PIDN:AAA65582.1; PID:9292836
A:Note: Authors translated the codon AGG for residue 1714 as Pro
C:Comment: Trichohyalin is a protein of the medulla of the hair and of the inner root sh
Covalent modifications to this protein include conversion of arginine to citrulline and

C:Genetics:
A:Gene: GDB:THH
A:Cross-references: GDB:I136223; OMIM:190370
A:Map position: lq21-lq21
C:Superfamily: trichohyalin; calmodulin repeat homology
C:Keywords: calcium binding; citrulline; EF hand; hair; tandem repeat
F:49-81/Domain: calmodulin repeat homology <BF2>
Query Match 3.6%; Score 310; DB 1; Length 1898;
Best Local Similarity 20.2%; Pred. No. 7.6e-06;
Matches 283; Conservative 203; Mismatches 459; Indels 454; Gaps 60;
Qy 350 ELHESAIYKATQISRRKHLFSRDK-----LKLFLKQHCPEQGVIKIKASSLSYKI 401
Db 484 ERHEQ-----ERRQQLKQDEEERRERWLKLEBEREQEER----- 521
Qy 402 AEQFSYFFDDPTTFISPANRRGRPPPKRIHIHSQEDNVANKOTLAS---YRSKATYER 458
Db 522 REQQ-----LRREQERERQRLK-----QEEERLQORLHSEQQLRREQEERL 565
Qy 459 DKLKQEMKSLAFK--AKLKR--EKADALEAKKKEKDEKKEKREELKKIVEERLKK 513
Db 566 EQLKREKLEQERREQLRQERDRKFRREEEQLLQEREQLRQERDRKFRREEEQLLQEREEL--EQLR-K 622
Qy 514 KEERERLKVREKEREKLR-----EEKRYVEYLQWQSKPREDMCEDDLKELPEPTPVKTR 569
Db 623 REEVERLEQE--ERRDERLKRPEEERRH-ELLK--SEEQERRHEQLRR--EQQERREQ 677
Qy 570 LPPPIFGDALMVLBNAGELFDLQDFPDGVTLEVLAEALVGNDSGGLCELLFPFLT 629
Db 678 LKRE-----EBE-----ERLEQLKREHEE-----E 698
Qy 630 AIFQAIAREEEVAKQQLTDADTKGCSLKSLDLSCTLSLRLHLILASGADVTSANAKY 689
Db 699 RREQELAEEOEQAREIKSRIPKW-----OWQLESEADARQSKVLL 740
Qy 690 RYQKRGGFDAITDQACMELRLSNPSLVKLSSTVYDLTPGKMKILHALCGKLLTLVSTR 749
Db 741 EAPQAGRAEAPQ--QEKRRRESELQWQEEERAHQEQEER-----R 783
Qy 750 DF-----IEDYVDILRQ-----AKQEFRELKAAQHRK-----EREAAARIRK 789
Db 784 DFTWQAEKESGRQRLSARPLRQERQRLAEERQERQERFLPEEKEQGRQR 843
Qy 790 E-----EKLKEQFQKMKQEKLEKDEQNSTADISIGEEERDF 829
Db 844 ERKELQLEEEQLRQERRAQQLQEEEDGLQEQERRRQERDRKFRREEEQLLQEREEL 902
Qy 830 DTSIESKDTQEKEL--DQDMFTEDDDPGSHKRRGRKRGQNGKFEFTROQINCVREL 887
Db 903 RHTLYAKPALQEQQLKQEQQLLQEEELQEREKRRRQEQE--RQYREEQLQEQEEL 960
Qy 888 LTADDEEALQKHQ--RKEKELLEKTSQAIACINIFPLG-----RDRMYRWY 934
Db 961 LREERKRRRQERERQYRKDKKQKEQL-----LGEPEKRRRQERREKRYR----- 1008
Qy 935 FPSIPGLFIEDYSLGTEDMLPPSPFQNNVQSDPOVSTK-----TGEPLMSESTNI 989
Db 1009 -----EEELQEEEQQLL--REERKRRRQERQYRKDEQLQEEELQEREEL 1059
Qy 990 DQGRDHSVOLPKPVHKPNRWCF-----YSSCEQLDQLIEAL----- 1026
Db 1059 RQERDRYREEELQEEEQEQQLLQEREETRREERTRRQERQYRKDEQLQEEELQEREEL 1118
Qy 1027 -NSRGHRESALKETLLQEKSRIC-----AQAR-FSEE-----KFHSDKPY- 1066
Db 1119 RQERERQEREELQEEEQEQQLLREERKRRRQERQYREEELQEREKRRQYRDEQ 1178
Qy 1067 -----PDSK-----PTYSRGRSSNAY---DPSQMAEKQLELRDLFLDIEDRIY 1109
Db 1179 SDLKQWQEFKENVAVNRKVKYCKRENEBQFLQSDSQV--RORQSQDQLHLQEQEER-- 1235
Qy 1110 QGTGLGAIKVTDRHIWRSR-----LESGRYELLSEBENKENGIIKTNEQV 1154

Db 1236 -----DREQERRRQOANRHPPEBEOLEBEQKEAKRDRKSQBEQO-----LUREERE 1284
Qy 1155 EMEIDEOT--KVIVDRLLIGIKTETPSTVSTNASTPQSVSVVHYLAMALFOIEQGIERR 1212
Db 1285 EKRRQETDRKPREEQQLQEREEO-----LLRQERDRK 1319
Qy 1213 FLKAPLDASDGRSVKTVLDRWRBSLLSASLSQVFLHLSTLDRSVIWSKSTLNARCKIC 1272
Db 1320 FREEELHQEQGRKFELEEQLRE-----ERBRKFLKEEQQLRLB-- 1359
Qy 1273 RKGDAENNVLCDCGDRGHHTYCVRPKLTVPEGDFCPECHPKORCRLSPRQPSLES 1332
Db 1360 -----ERQLQDRDRKPREEQQLS 1380
Qy 1333 DEDVDSMGGEVDGDEEQSEEBEYEBEQDDSDQSEEBEYELPKRGRPQVRLPVKT 1392
Db 1381 ROERDRKPREEQQVRRQERERKFLERQERHRKFREREQQLQEREEOQLRQERD 1440
Qy 1393 RGLKSSFSRQOQPGRYPSRSQOSTPKTTVSSKTGRSLR--KINSAPPTETKSLRIA 1450
Db 1441 RKFLERQQLRQERD-----RKFRQELRSQEP-ERKFLER 1477
Qy 1451 SRSTRHSHGLOADVVELLSPRKRGRKSNANTPENSPPNPNFVATKSSSEOSRSVN 1510
Db 1478 QQLHRQOR--QRKFLQEQQLRQERQORRQDRD-----KFR-----EEQLRQER 1523
Qy 1511 IASKLSQESBSKRCR-----KRQSP-----PSPVTLGRSSRQGGVHLSAF---EQLV 1560
Db 1524 EQQLSRQERDRKFLERQVRRQERQERKFMEDQQLRQEQOQLRQERDRKFRQERD 1583
Qy 1561 VE-----LVRHDDSMWFLK 1574
Db 1584 QERERQQLRQERDRKFLER 1602

RESULT 10
A:1928
cag island protein - Helicobacter pylori (strain J99)
C:Species: Helicobacter pylori
A:Variety: strain J99
C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 09-Jul-2004
C:Accession: A71928
R:Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.;
Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.B.; Vovis, G.F.;
Nature 397, 176-180, 1999
A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric path
A:Reference number: A71800; MUID:99120557; PMID:9923682
A:Accession: A71928
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1819 <ARN>
A:Cross-references: UNIPROT:Q9ZLV0; GB:AE001481; GB:AE001439; NID:94155005; PIDN:AAD0604
A:Experimental source: strain J99
C:Genetics:
A:Gene: orf13/14

Query Match 3.6%; Score 309.5; DB 2; Length 1819;
Best Local Similarity 19.4%; Pred No. 7.6e-06;
Matches 277; Conservative 225; Mismatches 520; Indels 407; Gaps 62;

Qy 312 ITSDSDSETQSCSFQNGKKKDAIDPL-----LFKYKVQPTKKELHESAIV 357
Db 212 ITDSDNDQEI---IKGSKKYIIGGIVAVLIVILFSLRSIFHYFV-PLD----- 258
Qy 358 KATQISRKRHLFSRDKLKL-----FLKHCEPQBGVTKIKASSLSTYKIAEQFSYFPF 411
Db 259 KSSRFKSDRNLYVNDIEIQIROYNRLKNERE-----KGNMI-----DKNLFFN 302
Qy 412 DDPPTFFIFSPAN--RRGRPPKRIHISOEDVAN-----KQTLASY-- 450
Db 303 DDPNRTLYNLIATIEDKNPLRPFYECISNGNGYECLEKLIKOKKLDQDMKKTLEAYND 362

Qy 451 ---RSKATYKRDK---LLKQEBMKSLEPEKAKLKRKAD-ALBAKKKSKEDKKEKREELK 503
Db 363 CTKNAKTEERIICLDLTKDENLK-----KSLNQKVQVALDCLKNAKTDEERK--ECL 415
Qy 504 KIVEERLKKKEKE-----RLKVEREK-----ERKLEBEKKYVEY 541
Db 416 KLINDPFIREFRKELELQKLEQYKOCIKNAKTEAKNECULGUSKGAERLKQOALDC 475
Qy 542 LKQMSKPREDECDLDELPEPTPVKTRLPPEIFGDALMVLEFLNAGELFLOQEFPPDG 601
Db 476 LKNAKTDEERKEC--LKNIPQ-----DLQKELLAD 503
Qy 602 VTLEVEBAL--VGNDSGPLCELLFFLTAIFAIAEBEE--EVAKEQLTDADTTGCSLK 658
Db 504 MSVKAYKDCVGRARNEKEKECEKL---LTPEAKKLLLENQALDCLKNAKTDEERKEC-LK 559
Qy 659 SLDDSDCTSLRIHLILASGADVTANAKYOKRGGFDATDDACMELRLSNPSLVKKL 718
Db 560 NLPKD--LQSDILAKESLKAYDCAS--QAKTEAEKK-----ECEKL----- 597
Qy 719 SSTSVYDLTPGCKMKILHALCGKLLTLVSTRDFIEDYVDILRQAQKQEPRELKAE-----Q 773
Db 598 -----LTP-EAKLLEE-----EAKESVKAYLDCVSAQKTEAEKECEKLLTPE 640
Qy 774 HRKREEAARIR-----KRKEEKLEQEQKMKKEKLEKEDQERNSTADISIGEE 825
Db 641 AKKLEEAKKSVRAYLDCVSKAKNAEAEKECEKLLTPEAKKLLLENQALDCLKNAKTDEER 700
Qy 826 RE-----DPDTIESKDTQEKELDQDMFTDEDD-----PGSHKGRGRKRGON 869
Db 701 KECLKDLPKDLQKVLAKESVRVYLDVCSKAKNAEAEKECEKLLTPEARKLLEAKKSVK 760
Qy 870 GFKEF---TROQINCVTRELLTADAEAL-----KOEHORKE----- 904
Db 761 AYKDCVLRARNEKEKECEKLLTPEARKLLEESKSVKAYLDCVSKAKNAEAEKECEKLL 820
Qy 905 ---KELLEKTQSAIA---CTNIFPLGRDRMYRRIYFIPSPIGLFIEDYSGLTEDML 956
Db 821 TPEARKLLEAKESVKAYKDCVSRARNEKEKECEKLLTPEARKLLEESKKS----- 872
Qy 957 PRPSFQNNVOSQPOVSTKTGEPLMSESTSNIDQPRDHSVOLPKPVHKPNRWCIFYSSC 1016
Db 873 --VKAYLDCVSKAKNAEAEKECEKLLTPEARKLLEAKESVKAYKDCVSRARNEKEKECE 930
Qy 1017 EQL-----DQLEAL-NSRGHRE-----SALKETLLOESRI---CAOLA 1052
Db 931 EKLITPEAKKLLLENQALDCLKNAKTEAEKCKVDLPKDLQKVLAKESVRVYLDVCSKA 990
Qy 1053 RFSEKHFHSDKPOPSKPTYSRGRSS--NAYDPSQMAEKQLELRDLFLDIEDR--Y 1109
Db 991 KNEAEKECEKLLTPEARKLLEAKESVKAYKDCVSRARNEKEKECEKLLTPEARKLLE 1050
Qy 1110 QGTIGAIVTDRHWRSALESGRYE---LLSEENK---ENGII-----KTWNED---VE 1154
Db 1051 QEVKSVKAYLDCVSRARNEKEKECEKLLTPEARKLLEENQALDCLKNAKTEAEKCKV 1110
Qy 1155 EMEIDEOTKVIVKDRLLGI-----KTETPSTVSTNASTPQSVSVVHY 1197
Db 1111 DLPKDLQKVLAKESVKAYLDCVSRARNEKEKECEKLLTPEARKLLEESKSVKAYLDC 1170
Qy 1198 LAMALFOIEQGIERRFL-----KAPLDASDSGRSYKTVLDRWR-----ESLSSASL 1244
Db 1171 VSKAKNAEKECEKLLTPEARKLLEAKESVKAYKDCVSRARNEKEKECEKLLTPEA- 1229
Qy 1245 SQVFLHLSTLDRSVIWSKSTLNARCKICRKKGDAENMVLCDCGDRGHHTYCVRPKLKTV 1304
Db 1230 -----RKLEQEY--KKSVKAYLDCVSRARNEKE----- 1256
Qy 1305 EGDWFCPECRKQCRRI-----SFRQPSLESDVEDSMGGEDDEVGDDEE----- 1352
Db 1257 -----KQCEKLLTPEARKFLERKQKQKDKAIKCLKNAD-----PNDRRAIMKCL 1302
Qy 1353 EQGSEEBEYEQDDEDDSQSEEBEYSLPKRGRPQVRLPVKTRGKLSLSSFSRQOQEPGRY 1412

```
Db 1303 DGLSDEEKLKYLQ---EAREKAVLDCLKTARTD-----EERKCKONLYSDLIQIQ-NKK 1353
Qy 1413 PSRQSQTPKTVTSKTRSRKNSAPPTETKSLRTASRSTRSHGPL-----QADV 1466
Db 1354 AQNKQNLQSKTERLHQASECLDNLDD--PTDQEAIEQCLEGLSDSERALILGIRKQADEV 1411
Qy 1467 VELLSPRRKR--GRKSNANTPENSFPNFRVIATKSSEQSRSNVIAS 1513
Db 1412 DRIYSDLSRKTDFNMAAGVPLPMDPKNGGDIATINATVWDADKIAS 1460

RESULT 11
A:Title: parasite-infected erythrocyte surface antigen MESA - malaria parasite (Plasmodium
A:Species: Plasmodium falciparum
C:Date: 22-Apr-1993 #sequence_revision 18-Nov-1994 #text_change 09-Jul-2004
C:Accession: A45605; A54517
R:Coppel, R.L.
Mol. Biochem. Parasitol. 50, 335-347, 1992
A:Title: Repeat structures in a Plasmodium falciparum protein (MESA) that binds human er
A:Reference number: A45605; MUID:92158014; PMID:1741020
A:Accession: A45605
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1526 <COP>
A:Cross-references: UNIPROT:P08116
A:Experimental source: Papua New Guinean isolate FC27
A:Note: sequence extracted from NCBI backbone (NCBIN:83648, NCBI:P:83656)
R:Coppel, R.L.; Culvenor, J.G.; Bianco, A.E.; Crewther, P.E.; Stahl, H.D.; Brown, G.V.;
Mol. Biochem. Parasitol. 20, 265-277, 1986
A:Title: Variable antigen associated with the surface of erythrocytes infected with matu
A:Reference number: A54517; MUID:87014571; PMID:3531849
A:Accession: A54517
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 222-443, 'K', 445 <C02>
A:Cross-references: GB:M15319; NID:g160060; PID:g552170
C:Superfamily: ring-infected erythrocyte surface antigen; dnaJ amino-terminal homology
C:Keywords: surface antigen; tandem repeat

Query Match 3.6%; Score 309; DB 2; Length 1526;
Best Local Similarity 17.3%; Pred. No. 6.5e-06;
Matches 218; Conservative 231; Mismatches 436; Indels 376; Gaps 47;
Qy 300 NGHNSVDGETIIISDSDSTQSC--SFQNGKKKDAIDP-LLPKYKVPQTKKELHESAI 356
Db 369 NSNVTSESEETKLAKEEENEGELGENVNDGASENSEDPKLTEQSENGTKSESEETKD 428
Qy 357 VKATQISRRKHLFSRDKLKLQHCPEQGVIKIKASSLSTYKIAEQDFSYFFPDPPPT 416
Db 429 DKPENEKADNKKKSKK-----KGFQFMGCGNFCNKNITD-----DEET 473
Qy 417 FIFSPANRRRPPRKRIHISQEDNVANKQTILA-----SYRSKATKRDKLL----- 462
Db 474 LVVKDDAKKHKFLREANTEKNDEKKDKLLGEGDKEDVKEKNDEQDKVLGEGDKEDVK 533
Qy 463 -KQEMKSLAPEKAKLREKADALPAKKKEKEDK-----EKURELKKIVBEERLKKKEEK 517
Db 534 EKNDQK----DKVLGEGDKEDVKE-KNDGKKDVGIGSEKTKQKEIKKV-EKRVKKKCKK 587
Qy 518 ERLVVEREKER-----EKUREKRVKVEYLKQWSKPREDMCEDDLKELPEPT 564
Db 588 KVKKGKENTEGNDKVKGPBIIIEVKBEIKKQVE-----DGIKENTEG 633
Qy 565 PVKTRLPPEIFGDALMVLEFLNFAFGELFDLQDFPDGVTLEVLBEALVGNDS-EGPLCEL 623
Db 634 NDVKV-GPEIITE-----EVKEEIKKQVEDGIKENDTEGNDKVKGP---- 673
Qy 624 LFFFLTAIFQIAABEEVEAKEQVTDATKGCSLKSLDLDSDCTSEILRLHLASGADVT 683
Db 674 ----EITERSVKEIKKQVEGKENTEG-----NDKVKGPBII 709
```

```
Qy 684 SANAKVRYQKR--GGFDATDDACMELRSLNPVLKLSSTSVYDLTPCEKMKILH---AL 738
Db 710 TEEVKEEIKKQVBGVKENDTESKDVIGQEIITEEVKKEIEKQEEKNGKENILEIKDIV 769
Qy 739 CGKLLTVSTRDFTEDVVD-----ILROAKQBFRELKAOHRKERE 779
Db 770 IGOEVIIIEEVKVKIKKVEKGIKENHTESKDVIGQEIIVEEVKEEI-EKQVEGKIKEND 828
Qy 780 -BAAARI-----RKRKEEKLKEQEKMKKEK-QSKLKEDEDEORNSTAD 818
Db 829 TESKDKVIGBEVIGKGVNNEGPNKQVTKQEKVKEVKEVKVKVKVKRNNKRNKRD 888
Qy 819 ISIGEE-EREDFDTSIESKDT--EQKELDQDMFTED--EDDPGSHKRGRRKRGQNGPKE 873
Db 889 NVIGKEIMKED----VNEKDTANKQKEIEQEKKEEVEKVEKVEKVEKVEKVEKEE 944
Qy 874 FTQEOQNCVTRRELLTADSEALKQEHORKEKELEKIQSAIACTNIPPLGRDRMYRYW 933
Db 945 VKEKEEVK-EKEEVKEDTSKDKIEQEKKEEVEKVEKEDTENKDKVIGQ----- 996
Qy 934 IFPSIPGLFIEEDYSGLTDMLLPRSPSFQNNVQSQDPQVSTKTKGELPMSESTNIDQGP 993
Db 997 -----IIIEE-----IKKEVKRVRKRNKNKNENKDNVIV----QBINMEDVNEKDTAN 1040
Qy 994 RDHSVQLPKPVHKPNRWCFYSSCEQLDQLIEALNSRGHRESALKETLLQEKSRICAQLAR 1053
Db 1041 KDKVIEQEK-----EKEEVKE-----KSEVKEEVEKVEEVE----- 1072
Qy 1054 FSEKHFHSDKPODPSKPTYSRGR-----SSNAYDPQWCAEKQL--ELRLRDFLLDI 1104
Db 1073 --KEKEVKEEVEKEDTSRDNVIVQEIWMNEDVNEKDTESKDKMIGKEVIEEVKEEV 1130
Qy 1105 EDRIYQGTGLGAIKVTDRHWRSALESGR-----YELLSEE----- 1139
Db 1131 KGRVN-----KEVNKRVRNRNRNRKQDVIEQEIIVSEVNEKDTKNNDKKIKGRVKKP 1183
Qy 1140 ----NKENGIIKTVNEDVEBMEIDEQTKVIVKORLL-GIKTETPTSTVSTNASTPQSVSSV 1194
Db 1184 IDCKKEREVQEIWMNEDVNEKDTESKDKMIGKEVIEEVKEEVEKRVNK----- 1232
Qy 1195 VHYLAMALFOIEQIERFLKAPLADSDGRSYKTVLDRWRRESLLSSASLSQVFLHSLTL 1254
Db 1233 -----EVNKRVRNRNRK-----NERKDVIEQEIIVSEVNEKDT- 1265
Qy 1255 DRSVINSKSIILNARCKICRKGDANNVLCGDCGRGHHYCYVRPKLTVPGDFWFCPCR 1314
Db 1266 -----KNNDKK-----IGKRVK-----K 1278
Qy 1315 PKQRCRLSFRQPSLESDEDDVEDSMGDEDDVDGDEEQSGSEEEYEVEQDEDDSQEE 1374
Db 1279 PIDCKKEREVQSESESESESESESESESESESESESESESESESESESESESESESE 1338
Qy 1375 EVSLPKRGRPOVRLPVKTRKGLSSFSRQGOQEPGRYPQRSQOSTPKTTVSSKTGSLR 1434
Db 1339 E-----ESESESESESESESESESESESESESESESESESESESESESESESE 1377
Qy 1435 K 1435
Db 1378 K 1378

RESULT 12
T34081
hypothetical protein C02F12.7 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 09-Jul-2004
C:Accession: T34081
R:Miller, N.
submitted to the EMBL Data Library, November 1995
A:Description: The sequence of C. elegans cosmid C02F12.
A:Reference number: Z21473
A:Accession: T34081
A:Status: preliminary; translated from GB/EMBL/DBDJ
```


Db 235 TKDNKENSISDKEDHQOEMLK--TLDKKERQKQEKEMKEQEKIEKKKKQEKQEKKK 292
Qy 515 EEKERLKVRE-----KERELREKKRYVEYLKQSKPREDMECDLDELPEPTP 565
Db 293 QEKERKQEKQEKQEKEMKKQKIEKERKKQEE--KEKKKKXKDKENEETMQPDQTS 350
Qy 566 VKTRLPPFIQDALMVLFLNFAFGLFDL--QDRFPDGVTLVLEALVGNDSGGLCEL 623
Db 351 EETN-----NEIMVPLP-----SPLTDVTTPEEHKEGEHKE--EEHKEGEHKEGEHKE- 396
Qy 624 LFFFLTAIFAIAABEEVEAKEQLTDATKG---CSLSKSLDSCITSEILRLHILAS-- 678
Db 397 -----EEHKEEHEHKEEHSKSKGKDKGKDKGKHKKAKKEVKVKKHVVKNVI 447
Qy 679 -----GADVTSANAKRYQKRGGFDDTDDACMELR-----LSNP--SLVKLSSTSV 723
Db 448 EDEKQGVETIINLEDK-----EACEBQHITVESRPLSQPOCKLIDPEQLTL 494
Qy 724 YDLTPGK--MKILHALCGKL--LTLVSTRDFIDYVDILRQAKQEPRELKAEQHRKERE 779
Db 495 MDKSVKEKNLSIOELIGTTGRVNVPRRNHKK-----KWAKIEAEALQKQKGVDEE 549
Qy 780 EAAAIRKRKEEKLQEQSKMKEKQEKLEDEQRNSTADISIGREERDFDFTSIESKYTE 839
Db 550 D-----KKEESKEQESKEQDEEVEEDEEESKEEESKEEESKEEESKEEESKEE 604
Qy 840 --QKELDDMTEDDDPGSHKRGGRGKRGQNGKFEFTRQIQNCVTRRELLTADBEALK 897
Db 605 EDEDEEDDAEEDDAEEDDAEEDDAEEDDDDEEDDDDE-----EDEDEDE 654
Qy 898 QEHORKEKELEKIQ 912
Db 655 EEEEEEESEKKIK 669

RESULT 14
A26655
myosin heavy chain [similarity] - slime mold (Dictyostelium discoideum)
N:Contains: myosin ATPase (EC 3.6.4.1)
C:Species: Dictyostelium discoideum
C:Date: 05-Oct-1988 #sequence revision 05-Oct-1988 #text_change 09-Jul-2004
A:Accession: A26655; A24728; S00250
R:Warrick, H.M.; De Lozanne, A.; Levinwand, L.A.; Spudich, J.A.
Proc. Natl. Acad. Sci. U.S.A. 83, 9433-9437, 1986
A:Title: Conserved protein domains in a myosin heavy chain gene from Dictyostelium discoideum
A:Reference number: A26655; MUID:87092266; PMID:3540939
A:Accession: A26655
A:Molecule type: DNA
A:Residues: 1-2116 <WAR>
A:Cross-references: UNIPROT:P08799; GB:M14628; GB:M11938; NID:g167834; PIDN:AAA33227.1;
R:DeLozanne, A.; Lewis, M.; Spudich, J.A.; Levinwand, L.A.
Proc. Natl. Acad. Sci. U.S.A. 82, 6807-6810, 1985
A:Reference number: A24728; MUID:86016788; PMID:3901008
A:Accession: A24728
A:Molecule type: mRNA
A:Residues: 2035-2116
R:Waggle, G.; Noegel, A.; Scheel, J.; Gerisch, G.
FEBS Lett. 227, 71-75, 1988
A:Title: Phosphorylation of threonine residues on cloned fragments of the Dictyostelium myosin heavy chain
A:Reference number: S00250; MUID:89112226; PMID:2828113
A:Accession: S00250
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1734-1893 <WAG>
C:Comment: The rod domain is highly periodic, containing a pattern of 7-residue repeats
C:Superfamily: myosin heavy chain; myosin motor domain homology
C:Keywords: actin binding; ATP; coiled coil; hydrolase; nucleotide binding; P-loop; phosphatase
F:1-818/Domain: globular head <HED>
F:89-747/Domain: myosin motor domain homology <MMOT>
F:179-186/Region: nucleotide-binding motif A (P-loop)
F:819-2116/Domain: alpha-helical rod <ROD>

Query Match 3.5%; Score 298; DB 2; Length 2116;
Best Local Similarity 21.2%; Pred. No. 2.8e-05;
Matches 294; Conservative 205; Mismatches 519; Indels 372; Gaps 64;
Qy 315 DSDSETQSCSFQNGKKKDA-----IDPLLFKYKVOPTK-----KELHES 354
Db 715 DAEDSQ-----KATDAVLKHLNIDPEQYRFGI--TKIFFRAGOLARIEAREQRIS 763
Qy 355 AIVKATQISRKRHLFSRDKLKLFLKQHCPEQEGVGIKIKASSLSYKYKAEODFSYF--PPD 412
Db 764 EIIKAIQAATRGWTARK-----VYKQAREHTVAARIQOOLRAYIDPKS 807
Qy 413 DPPTFIFSPANRRGRP--PKRIHISQEDNVANKQ--TLASYRSKATKRDKL--LKQEE 466
Db 808 WPMWKLFSKA-----RPLKRNPEKEIKEREILELKSNTLDTSTQKOKLESLKQDTE 862
Qy 467 -----MKSIAFEKAKLKR--EKADALEAKKKE-----KEDKEKREELKKIIVEERLKKK 514
Db 863 SNVLQQLQKAEKETLKAAMYDSKDALEAKRELEIRVEDMESELEDEKCLALENLQNR 922
Qy 515 --BEKERLKVREKEREKLEEE---KRKYVEYLKQSKPREDMECDLDELPEPTPVT 568
Db 923 SVEEKVLDLEELQEEQKRLNTLEKKKKVEEELKEE--MKRVNDGQSDTISLEK--IKD 978
Qy 569 RLPPEIFGDALMVLFLNFAFGLFDLQDEFPDGVTL-----VLEEALVGNDSGGL 620
Db 979 ELQKE-----VVELTESFSE-----ESKDKGVLEKTRVRLQSELDDLTVRLDSETKD 1025
Qy 621 CELFFFLTAIFAIAABEEVEAKEQLTDATGTCGSKLSLDLSDCTISEILRLHILASG- 679
Db 1026 KSELLRQKKLEELKQVEALAAETAAKLAQAAN--KQLQGEYTELNEKFNSEVTARSN 1084
Qy 680 -----ADVTSANAKRYQKRGGFDDT-----DACMELRSLPSLVKLSST--- 721
Db 1085 VEKSKTLESQLVAVNNLEDEEKKNR--DALEKKKKALDAMLE-----EMKQLESTGGE 1137
Qy 722 --SVYDLTPGKMKILHALCGKLTLVSTRDFIDYVDILRQAKQEFR--ELKAEQHRKER 778
Db 1138 KKSILYDLKVQESD--MEALRNQISELQSTIAKLEIKSTLEGEVARLQGELEAEQLAKSN 1196
Qy 779 -----EAAAIRKRKEEKLKEQ--EQMKKEKQEKLEKEDSQRNSTA 817
Db 1197 VEKQKKVLDLEDKSAQLAEETAA--KQALDKKKLEQELSEVOTQLSEANNKNVNS 1253
Qy 818 DISIGSEEREDFTSIESKTEQKELQDMFTEDEDDPGSHKRGGRGKRGQNGKFEFTRQ 877
Db 1254 D--STNGHLETSFNNLKLLEAEAKA-----KQALEKKR-----LGLE 1289
Qy 878 EQINCVTRRELLTADBEALQEHORKEKELLEKLEKIQSAIACTNIPFLGRDRMYRRYIPPS 937
Db 1290 SELKHVNEQL-----EEKKQKESNEKRVLEKEVSEL-----KDQ----- 1326
Qy 938 IPGLFIEEDYS--GLTEDMLLPPRSPFNQVNOQDPQVSTKTGCEPLMSESTSNIDQGR 994
Db 1327 -----IEEVASKKAVTE-----AKNKESELDEIKRQYADVVS-----R 1362
Qy 995 DHSVOLPKPVHKPNRCWCFYSSCEQDOLJIEALNSRGHRESALKETL--LOEKS--RICAQ 1050
Db 1363 DKSVEQLTKLQAKNEELRNLTAEABGQLDRAERSKKKAEFDLEAVKNLEEBETAKVKAE 1422
Qy 1051 LARFSEKHFHSDKPDSPDKPTYSRGRSSNAYDPSQCAEKQLRLRLRDLFLDIEDRIYQ 1110
Db 1423 KAMKKAETDYRSTKSELDD-----AKNVSSQYVQIKRLNEELSELR---SVLEADERCN 1475
Qy 1111 GTLGAIKVTRDHWRSALESGRVELLSEEN--KENGIIKTVNEDVEEIDBQTK----- 1163
Db 1476 SAIKAKKTAE-----SALESILKDEIDAANNAKAAERKSELEVRVAELESLEDKSGTV 1530
Qy 1164 -----VIVKRLILGITEPSTYSTVSTNASTPQSVSSVVHYLAMALFOIEQGIERRFL 1214
Db 1531 NVEFIRKQDAIBDLRLARDRETESRKSD-----EDKQNKIKQFA 1571
Qy 1215 KAPLDASDGRSYKTVLDRWRRESLSSASLSQVFLHLST--LDRSVIWSKSLNARCKICR 1273

```

Db 1572 DLEAKVEAQREVVTI-DRLKKKLESD-----IIDLSQLDTER-----KSRIKIEK 1617
Qy 1274 KKGDAENMYLDCDGRGHHTYCVRPKLTVPBGDWFCEPCRPKQRCRLSFRQPSLRSD 1333
Db 1618 SKKLE-----QTAE-----RRAAEGSSKAAD 1641
Qy 1334 EDVESMGEDDEVG--DEBQSQEHEEYEV-----RODEDSQEEBEV-----SLPKRG 1382
Db 1642 EIRKQVQVQVDELDAQLDSEAAALNASEKKIKSLVAEYDEVEKEQLEDEILAKDLVAK 1701
Qy 1383 RP-QVRLPVKTRGKLSFSSFGQQQEPGR-----YPSRSQOSTPKTVSSKT 1429
Db 1702 RALEVELE-EVRDQLEEDSRSELEDSKRRTTTEVEDIKKYDAEVEQNTKLDAAKKL 1760
Qy 1430 GRSRLKINSAPPTETKSLRIASRTRSHGPGLOADVVELLSPRKRGRKRSANNTP-EN 1488
Db 1761 TDDVTLKKQLEDEKKLNESERAKKLES--ENEDFLAKDAEYKNSRAEKORKKYEK 1818
Qy 1489 SPNFPNFRV---IATKSSEQRSVNIAKLSIQESKRRCKRKQSPSPVTLGRRSSG 1545
Db 1819 DLKDTKYKLNDEAATKTQTEGAACKLEQIDELRSKLEQEQAKATQADSKKTL----- 1872
Qy 1546 ROGGVHELSEA 1555
Db 1873 -EGEIDNLRA 1881

RESULT 15
A42184
nuclear mitotic apparatus protein NUMA - human
C:Species: Homo sapiens (man)
C:Date: 31-Dec-1993 #sequence revision 31-Dec-1993 #text_change 09-Jul-2004
C:Accession: A42184; S33376; S55331; S23647; S24554
J:Compton, D.A.; Szilak, I.; Cleveland, D.W.
J:Cell Biol. 116, 1395-1408, 1992
A:Title: Primary structure of NUMA, an intranuclear protein that defines a novel pathway
A:Reference number: A42184; MUID:92176238; PMID:1541636
A:Accession: A42184
A:Molecule type: mRNA
A:Residues: 1-2101 <COM>
A:Cross-references: UNIPROT:Q14981; EMBL:Z11584; NID:G35120; PIDN:CAA77670.1; PID:G35121
A:Note: Sequence extracted from NCBI backbone (NCBIN:85755, NCBIP:85760)
R:Tang, T.K.; Tang, C.C.; Chen, Y.L.; Wu, C.W.
J:Cell Sci. 104, 249-260, 1993
A:Title: Nuclear proteins of the bovine esophageal epithelium. II. The NuMA gene gives a
A:Reference number: S33376; MUID:93280231; PMID:8505359
A:Accession: S33376
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1705-2101 <TAN>
A:Cross-references: EMBL:Z14229; NID:G296118
A:Note: This translation is not annotated in GenBank entry HSNUMAT3G, release 113.0
R:Harborth, J.; Weber, K.; Osborn, M.
EMBO J. 14, 2447-2460, 1995
A:Title: Epitope mapping and direct visualization of the parallel, in-register arrangement
A:Reference number: S55331; MUID:95300777; PMID:7781599
A:Accession: S55331
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 247-279 <HAR>
R:Yang, C.H.; Lambie, E.J.; Snyder, M.
J:Cell Biol. 116, 1303-1317, 1992
A:Title: NuMA: an unusually long coiled-coil related protein in the mammalian nucleus.
A:Reference number: S23647; MUID:92176231; PMID:1541630
A:Accession: S23647
A:Molecule type: mRNA
A:Residues: 1-771, 'Q', 773-814, 'BR', 817-872, 'E', 874-1267, 'RLRLQAETAGNSARAERSALREEVQSLR
A:Cross-references: EMBL:Z11583; NID:G35118; PIDN:CAA77669.1; PID:G35119
A:Note: the authors translated the codon GAG for residue 781 as Gly, TTC for residue 176
C:Genetics:
A:Gene: GDB:NUMA1; NUMA

```

```

A:Cross-references: GDB:137229; OMIM:164009
A:Map position: 11q13-11q13
C:Keywords: mitosis; nucleus

Query Match 3.4%; Score 297; DB 2; Length 2101;
Best Local Similarity 20.0%; Pred. No. 3.1e-05;
Matches 271; Conservative 210; Mismatches 486; Indels 388; Gaps 61;

Qy 340 FKYKVPQTKGLHSAIVKATQISRRKHLFRDRLKFLKQKCEPQEGVIRIKASSLSY 399
Db 119 FEYKIQ-----ELAVILKFVLDHEDGLNLEDENFLQ-----KAPVPSI- 159
Qy 400 KIAEQDFYFPDPPPTPIFSPANRRRRPKRIHISQEDNVNANKQTTLASVRSKATKRD 459
Db 160 -----CSSTPPEE-----LSPPSHQAAREIRFLELQKVASSSSNNFLS-GSPASPMGD 207
Qy 460 KLLKQEMKSLAFKAKLREKADALEAKKKEKEDKREBELKKIVE-----508
Db 208 IL-----QTQFQNRRLUKQLAD-----ERSNRDELELELAENRKLKTEKQAIANMOOR 257
Qy 509 -ERLKKKEERLKYVEREKREKREKRYVEYLKQWSPREDMECDLDELPEPTPVK 567
Db 258 IDRLALLNEKQAASPLEPKLEELRDKNESLTMRHLTK-----QCQDLKTEKQMDRK 312
Qy 568 TRLPPEITGD-ALMVLEFLNAGBELFDLQDFFPD---GVTLVLE-ELAVGNDSGLPCE 622
Db 313 INQLSENGDLSFKLREFASHLQQLDALNELTEHSHKATQEWLEKQAKLEKLSAALQD 372
Qy 623 LLFFFLTAIFOAIAEEEEVAKEQITDADTKGCSLKSLDLDSCTLSEILRLHLTASGA 680
Db 373 -----KKCLEKNEILQGLKSLDEHLSQLQDNPPOEKGVEVLGDVQLTLEKQEA 422
Qy 681 DVTGAN-----AKYRYQKRGGFATDDACMELRLNSLSLVKLSST 721
Db 423 ATLAANTQLQARVEMLETERGQEQEAKL-LAERGHFE-----BEKQQLSSLITDLQS- 473
Qy 722 SVYDLTPG-----EKMKILH-----ALCGKLLTLVSTRDIEDYVDILRQAKQEPRE 768
Db 474 SISNLSQAKERLEQASQAAGARLTAQVASLTSELTTLNAT---IQQQQELAGLQQAQKE 530
Qy 769 LKAE--QHRKEREAAAIRKKEE---KLKEQKQKKEKQKLEKEDQORNSTADISIGE 823
Db 531 KQAQLAOTLQOQEQASQGLRHQVQLSSLSKQKQKLEKVAEK-QSATQDHAQOALATAA 589
Qy 824 EEREDFTDTSIESKQTEQKELQDQDMFTDEDDPGSHKRGKRGKQNGKFEFTROQINCVC 883
Db 590 EERE---ASLREDAALKQLEA-----LEKEKAALKLEIL 620
Qy 884 TRELLTADDEEALKQ-----EHQRKEKELLEKIQSAIACTNIFPLGRDRMYRRIYFPISIP 939
Db 621 QOQLQVANEARDSAQTSTVQAQREKAEKLSRKEVEELQAC-----658
Qy 940 GLFTIEDYSGITEDMLPRSPSFQNNVQSQDPQVSTKTGEPLMSESTSNIDQGPDRHSVQ 999
Db 659 -----VETARQEQHEAQAQVAEELQ-----LRSEQ-----684
Qy 1000 LPKPVHPNRMCFYSSCEQLDQLTALNSRGHRESALKETL-----LOEKSRIACAQLAR 1053
Db 685 -QKATEKER-----VAQEKDQLQEQ-----QALKESLKVTKGSLKEEKRAADAL- 729
Qy 1054 FSEKPHFSDKPDPSKPTYSYRGRSSNAYDPSQMAEQKLELRDRDF--LLDIEDRIYQG 1111
Db 730 --EQQRCSISELKAEATRSVLQVHQHREKRELEERAGRGLEARLLQLGEAQAETEVLR 787
Qy 1112 TLGAIK-----VTDRHIWR-----SALESQRYELLSEENKENGIIKTVNEDV 1153
Db 788 ELAEAMAAQHTAESECEQLVKEVAAWRDGVDSDSQEEAQYGAAMFQEQ-----LMTLKSEC 842
Qy 1154 EE--WEIDEQTKVIVKDRLLIGIKTETPTSTVTNASTPOSVSVVH-YLAMALFOIEQOIE 1210
Db 843 EKARQEIQE-----AKKVAGIESHSELQISRQ-----QNKLAELHANLALALQVQVE--- 890
Qy 1211 RRFLKAPLDASDSGRSYNTVLDRWRRESLLSSASLSQVFLHLSTL-DRSVIWSKSLNARC 1269

```


Db	891	-KEVRAQKLAADD-----LSTLOEKMAATSKEV--ARL 919
Qy	1270	K-ICRKKGDAAENVLCDCGRGHHTYCVRPKLKTVPPEGDWFCPCPKORCRLRSFRQP 1328
Db	920	ETIVRKAGEQDET-----ASRELVEKPAPAGDRQPEWLEEQGRCFTQ-A 965
Qy	1329	SLESDEVEDSMGGEDDEVGD--EBEGQSEEEYEVEQDDSDQBEEEVS--LPKRGRP 1384
Db	966	ALQAMEREAEQMGNELERFAALMESQGQQOEE-----GQOREVARLTQERGA 1016
Qy	1385	QVLPLVK--TRGKLSSSFSGRGQQBPGRYPSPSQOSTPKNVTSSKTKGR-----SLRKIN 1437
Db	1017	QADLALKAAREALEMRLQNALNEQ--RVFATLQEALAHALTKEKGDKQLAKLRGLE 1073
Qy	1438	SAPTETKSIR-----IASRSTRHSHGPLOADVVELLSPPRRKRRGRKS--ANNTP 1486
Db	1074	AAQIKLEELURQTQVKLQELAKKXEHASG-----SQAQSEAAGRTE 1116
Qy	1487	ENSPNPNFNPIATKSEQ-----SRSVNTASKLSLOESBKRRCKRQSPSPS 1535
Db	1117	PTGPKLEALRAEVSKLEQQCKQEQEADSLSRLSLE-AERASRAERDSALLETLOGGLEKA 1175
Qy	1536	PVTIGRSSRGOGGVHELSAFEOLVVVELVRHDDSW 1570
Db	1176	Q-ELGHSQSALASAQRELAAFRTKVQDHSKAEDW 1209

RESULT 16

S38173
myosin-like protein MLP1 - Yeast (*Saccharomyces cerevisiae*)
N;Alternate names: protein YKR095w; protein YKR415
C;Species: *Saccharomyces cerevisiae*
C;Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 09-Jul-2004
C;Accession: S38173; S40647; S31207
R;Baladron, V.; Ballesta, J.P.G.; Bou, G.; del Rey, F.; Esteban, P.F.; Garcia-Cantalejo, M.; Genet, N.; Nguyen, E.Y.; Botstein, D.
submitted to the Protein Sequence Database, March 1994
A;Reference number: S38158
A;Accession: S38173
A;Molecule type: DNA
A;Residues: 1-1875 <BAL>
A;Cross-references: EMBL:X73541; NID:g450550; PIDN:CAA51948.1; PID:g450554
A;Experimental source: strain S288C
R;Koelling, R.; Nguyen, T.; Chen, E.Y.; Botstein, D.
Mol. Gen. Genet. 237, 359-369, 1993
A;Title: A new yeast gene with a myosin-like heptad repeat structure.
A;Reference number: S31207; MUID:93247549; PMID:8483450
A;Accession: S31207
A;Molecule type: DNA
A;Residues: 1-1875 <BOU>
A;Cross-references: EMBL:X73541; NID:g450550; PIDN:CAA51948.1; PID:g450554
A;Experimental source: strain S288C
R;Koelling, R.; Nguyen, T.; Chen, E.Y.; Botstein, D.
Mol. Gen. Genet. 237, 359-369, 1993
A;Title: A new yeast gene with a myosin-like heptad repeat structure.
A;Reference number: S31207; MUID:93247549; PMID:8483450
A;Accession: S40647
A;Molecule type: DNA
A;Residues: 1-1875 <BOU>
A;Cross-references: EMBL:X73541; NID:g450550; PIDN:CAA51948.1; PID:g450554
A;Experimental source: strain S288C
R;Koelling, R.; Nguyen, T.; Chen, E.Y.; Botstein, D.
Mol. Gen. Genet. 237, 359-369, 1993
A;Title: A new yeast gene with a myosin-like heptad repeat structure.
A;Reference number: S31207; MUID:93247549; PMID:8483450
A;Accession: S40647
A;Molecule type: DNA
A;Residues: 1-1875 <BOU>
A;Cross-references: EMBL:X73541; NID:g450550; PIDN:CAA51948.1; PID:g450554
A;Experimental source: strain S288C
R;Koelling, R.; Nguyen, T.; Chen, E.Y.; Botstein, D.
Mol. Gen. Genet. 237, 359-369, 1993
A;Title: A new yeast gene with a myosin-like heptad repeat structure.
A;Reference number: S31207; MUID:93247549; PMID:8483450
A;Accession: S40647
A;Molecule type: DNA
A;Residues: 1-1875 <BOU>
A;Cross-references: EMBL:X73541; NID:g450550; PIDN:CAA51948.1; PID:g450554
A;Experimental source: strain S288C
R;Koelling, R.; Nguyen, T.; Chen, E.Y.; Botstein, D.
Mol. Gen. Genet. 237, 359-369, 1993
A;Title: A new yeast gene with a myosin-like heptad repeat structure.
A;Reference number: S31207; MUID:93247549; PMID:8483450
A;Accession: S40647
A;Molecule type: DNA
A;Residues: 1-1875 <BOU>
A;Cross-references: EMBL:X73541; NID:g450550; PIDN:CAA51948.1; PID:g450554
A;Experimental source: strain S288C
R;Koelling, R.; Nguyen, T.; Chen, E.Y.; Botstein, D.
Mol. Gen. Genet. 237, 359-369, 1993
A;Title: A new yeast gene with a myosin-like heptad repeat structure.
A;Reference number: S31207; MUID:93247549; PMID:8483450
A;Accession: S40647
A;Molecule type: DNA
A;Residues: 1-1875 <BOU>
A;Cross-references: EMBL:X73541; NID:g450550; PIDN:CAA51948.1; PID:g450554
A;Experimental source: strain S288C
R;Koelling, R.; Nguyen, T.; Chen, E.Y.; Botstein, D.
Mol. Gen. Genet. 237, 359-369, 1993
A;Title: A new yeast gene with a myosin-like heptad repeat structure.
A;Reference number: S31207; MUID:93247549; PMID:8483450
A;Accession: S40647
A;Molecule type: DNA
A;Residues: 1-1875 <BOU>
A;Cross-references: EMBL:X73541; NID:g450550; PIDN:CAA51948.1; PID:g450554
A;Experimental source: strain S288C
R;Koelling, R.; Nguyen, T.; Chen, E.Y.; Botstein, D.
Mol. Gen. Genet. 237, 359-369, 1993
A;Title: A new yeast gene with a myosin-like heptad repeat structure.
A;Reference number: S31207; MUID:93247549; PMID:8483450
A;Accession: S40647
A;Molecule type: DNA
A;Residues: 1-1875 <BOU>
A;Cross-references: EMBL:X73541; NID:g450550; PIDN:CAA51948.1; PID:g450554
A;Experimental source: strain S288C
R;Koelling, R.; Nguyen, T.; Chen, E.Y.; Botstein, D.
Mol. Gen. Genet. 237, 359-369, 1993
A;Title: A new yeast gene with a myosin-like heptad repeat structure.
A;Reference number: S31207; MUID:93247549; PMID:8483450
A;Accession: S40647
A;Molecule type: DNA
A;Residues: 1-1875 <BOU>
A;Cross-references: EMBL:X73541; NID:g450550; PIDN:CAA51948.1; PID:g450554
A;Experimental source: strain S288C
R;Koelling, R.; Nguyen, T.; Chen, E.Y.; Botstein, D.
Mol. Gen. Genet. 237, 359-369, 1993
A;Title: A new yeast gene with a myosin-like heptad repeat structure.
A;Reference number: S31207; MUID:93247549; PMID:8483450
A;Accession: S40647
A;Molecule type: DNA
A;Residues: 1-1875 <BOU>
A;Cross-references: EMBL:X73541; NID:g450550; PIDN:CAA51948.1; PID:g450554
A;Experimental source: strain S288C
R;Koelling, R.; Nguyen, T.; Chen, E.Y.; Botstein, D.
Mol. Gen. Genet. 237, 359-369, 1993
A;Title: A new yeast gene with a myosin-like heptad repeat structure.
A;Reference number: S31207; MUID:93247549; PMID:8483450
A;Accession: S40647
A;Molecule type: DNA
A;Residues: 1-1875 <BOU>
A;Cross-references: EMBL:X73541; NID:g450550; PIDN:CAA51948.1; PID:g450554
A;Experimental source: strain S288C
R;Koelling, R.; Nguyen, T.; Chen, E.Y.; Botstein, D.
Mol. Gen. Genet. 237, 359-369, 1993
A;Title: A new yeast gene with a myosin-like heptad repeat structure.
A;Reference number: S31207; MUID:93247549; PMID:8483450
A;Accession: S40647
A;Molecule type: DNA
A;Residues: 1-1875 <BOU>
A;Cross-references: EMBL:X73541; NID:g450550; PIDN:CAA51948.1; PID:g450554
A;Experimental source: strain S288C
R;Koelling, R.; Nguyen, T.; Chen, E.Y.; Botstein, D.
Mol. Gen. Genet. 237, 359-369, 1993
A;Title: A new yeast gene with a myosin-like heptad repeat structure.
A;Reference number: S31207; MUID:93247549; PMID:8483450
A;Accession: S40647
A;Molecule type: DNA
A;Residues: 1-1875 <BOU>
A;Cross-references: EMBL:X73541; NID:g450550; PIDN:CAA51948.1; PID:g450554
A;Experimental source: strain S288C
R;Koelling, R.; Nguyen, T.; Chen, E.Y.; Botstein, D.
Mol. Gen. Genet. 237, 359-369, 1993
A;Title: A new yeast gene with a myosin-like heptad repeat structure.
A;Reference number: S31207; MUID:93247549; PMID:8483450
A;Accession: S40647
A;Molecule type: DNA
A;Residues: 1-1875 <BOU>
A;Cross-references: EMBL:X73541; NID:g450550; PIDN:CAA51948.1; PID:g450554
A;Experimental source: strain S288C
R;Koelling, R.; Nguyen, T.; Chen, E.Y.; Botstein, D.
Mol. Gen. Genet. 237,

Query Match	3.4%	Score 295.5;	DB 2;	Length 1875;
Best Local Similarity	18.4%;	Pred. No. 3.2e-05;		
Matches 317;	Conservative 270;	Mismatches 563;	Indels 575;	Gaps 75;
Qy	261	YVKDRIYVEETVEIRN-----NGARLQCTILEVLPSSHQGFANGHVNSVDGETII--	312	
Db	225	YLSVYRKTDKVIDIRNELNRLNDFPOWERNTNDVLKQK-----NNELSKLOEKLEII	278	

[illegible]

Db 1250 NVAPIESELTALKYMQKQKQELKLAKEVEHHRWKKRSQDILLEKHQOLSSDYKLESIE 1309
Qy 1154 ---EEMEIDEQTKVVKORLLGIKTETSTVSTNASTPQSVSSVHYL----- 1198
Db 1310 NLKELENKERGAEBKFNRLRQAOERLTKSLQSDSLTEQVNSLRDAKNVLENSLS 1369
Qy 1199 -----ANALFOIEQGTIER--RFLKAPLDASDSGRSKYTKVLDWRRE 1236
Db 1370 EANARIEBLONAKVAQGNQLEAIKRLQEDAEKASRELQAKLE--ESTSYESTINGLNE 1427
Qy 1237 SLLS-----SASLSQVFLHLSTLDRSVIWSKILNARCKICRKGDAENM 1281
Db 1428 EITTLKEIEIKORQIQOOLQATSANEQNDLSNIVESM--KKSFEEDKIKFIKEKTQEVNE 1485
Qy 1282 VLDCGCDR-----GHHTYCVRPKLKTVPEDWFCPCPRKQRCRRLSF 1324
Db 1486 KILEAQERLQNSINMEEIKKWESEHEQVQSQIREAEAEALKKIRLPTEKINKIIE 1545
Qy 1325 RQPSLES--DEDVED-----SMGDEDEVGDDEEGQSEBEEVEVEQEDDDSOBEEVS 1377
Db 1546 RKKEBEKEFEKVBKRIKMSQSGEIDVVLKQLQLEAKVQEKQKELENEYNKKLQELKD 1605
Qy 1378 LPKRG-----RPQVRLPVKTRGLSSSFSSRGQQQEPGRYPSPRSQQSTPKTIVSSKTCR 1431
Db 1606 VPHSHSIDDERDKLURABIESR--LREBFNNELOAIKKKSFDEGQQAAMKTTLE---R 1660
Qy 1432 SIRKINS-----APPTET-----KSLRIASR 1452
Db 1661 KLAKMESQLSETKQSAEPPKSVNNVQNPPLGLPKRIENSNSPNNPLLSGEKLLKNSK 1720
Qy 1453 STRHSHGFLQADVVELLSPRR-----KRGKRSANNTPE--ENSPNFPNFRVIATKS 1502
Db 1721 SSSGGFNFTSP-----SPNKHQNDNDKRESLANKTDPPTLEPSPFNIPASRGLISS 1774
Qy 1503 SEQSRSVN---IASKLSIQESKRCRCRKQSPPEP-----SPVTIG 1540
Db 1775 STLSTDTNDEELTSNPAQKSSNNRNVQSEEDTEKKKEGEPVKRG 1819

RESULT 17
A45592
liver stage antigen LSA-1 - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:Date: 22-Nov-1993 #sequence revision 02-Dec-1994 #text change 09-Jul-2004
C:Accession: S24597; A45592; S29393; S34842; B45592; C45592; D45592
R:Zhu, J.; Hollingdale, M.
submitted to the EMBL Data Library, November 1990
A:Reference number: S24597
A:Accession: S24597
A:Molecule type: DNA
A:Residues: 1-1909 <ZHU>
A:Cross-references: UNIPROT:Q25893; EMBL:X56203; NID:G9915; PID:G9916
R:Zhu, J.; Hollingdale, M.R.
Mol. Biochem. Parasitol. 48, 223-226, 1991
A:Title: Structure of Plasmodium falciparum liver stage antigen-1.
A:Reference number: A45592; MUID:92107224; PMID:1840628
A:Accession: A45592
A:Molecule type: DNA
A:Residues: 1-195:638-688; 1165-1215; 1590-1909 <ZH2>
A:Note: sequence extracted from NCBI backbone (NCBIN:75010, NCBIN:75014, NC
R:Guérin-Marchand, C.; Druilhe, P.; Galey, B.; Londono, A.; Patrapotikul, J.; Beaudoin,
Nature 329, 164-167, 1997
A:Title: A liver-stage-specific antigen of Plasmodium falciparum characterized by gene c
A:Reference number: S29393; MUID:87315391; PMID:3306406
A:Accession: S29393
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 323-387 <GUE1>
A:Cross-references: EMBL:M28266
R:Guérin-Marchand, C.; Druilhe, P.; Galey, B.; Londono, A.; Patrapotikul, J.; Beaudoin,
submitted to the EMBL Data Library, April 1992
A:Description: a liver-stage-sepcific antigen of plasmodium falciparum characterized by

A:Reference number: S34842
A:Accession: S34842
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 323-381, 'HKAI', 'GUE2>
A:Cross-references: EMBL:M28266
A:Note: difference at carboxyl end due to frameshift error
C:Comment: This protein is found as flocculent material in the parasitophorous vacuole.
C:Superfamily: trichohyalin; calmodulin repeat homology
C:Keywords: EF hand
F:154-1629/Region: 17-residue repeats (A-K-E-K-L-Q-E-Q-Q-Q-D-L-E-Q-E-Q-R-R)

Query Match 3.4%; Score 295; DB 2; Length 1909;
Best Local Similarity 18.1%; Pred. No. 3.4e-05;
Matches 234; Conservative 263; Mismatches 50; Indels 294; Gaps 50;

Qy 329 GKXDAIDPLLFKVKVQPTKKELHESATVK---ATQISRRKHLFSRDKLFLKQHC- 383
Db 364 QQQSDLEQERLAKELQEQSDLEQDRLAKELQEQSDLEQERLAKELQ---EQSD 420
Qy 384 -----POGVIKIKASSLSTYKIAEQDFSYFFPDPTTFIPSPANRRGRPKRIH 434
Db 421 EQERRAKEQEQSDLEQERLAKELQEQ-----SDLEQER-RAKELQ 465
Qy 435 ISQSD---NVANKOTLASYRSKATKER---DKLLKQE---EMKSLAFKAKLKRKADAL 485
Db 466 EQSDLEQERRAKEQEQSDLEQERLAKELQEQSDLEQERLA---KEKLEQEQSDS- 522
Qy 486 EAKKKEDEKKEEELKVIIEERLKKKE-----EKERLKVVEREKEREKREK 537
Db 523 ---EQERLAKELQEQ-OSDLEQERLAKELQEQSDLEQERLAKELQEQSDLEQERL 578
Qy 538 YVEYLKQWSKPREMECDLDELPEPTPVKTRLPPEIFGDALMVLEFLNAQELFDLOE 597
Db 579 AKELQEQ---QQSDLEQERL-----AKEKQ-----QQSDLEQ 610
Qy 598 FPDGVTLVEALVGNDSQGLCELLPFFLTAIQAIAEBEEVAKSOLTADTKGCSL 657
Db 611 -----RLAKEKQQQSD-----LEQERLAKELQEQSDS----- 639
Qy 658 KSLDLDSCITLSEILRLHLASGADVTSANAKYRQKRGFDATDDACMELRLSNFLVK 717
Db 640 ---DLEQERLAK-ELQEQSDLEQERLAKELQEQSDLE-----OERLAKELQEQ 688
Qy 718 LSSTSVVDLTGPKMKILHALCGKLLTLVSTRDFTEDYVILRQAKQFRELKAEQHRKE 777
Db 689 QS-----DLEQERAK-----EKLEQEQSDLEQERRAKEKLEQEQSDLE 727
Qy 778 REEAARIRKKEKELK-----EQQMKKEKQKELKEDEQNRNSTADISIGEEEREDFT 831
Db 728 QE-----RRAKEKLEQEQSDLEQERRAKEKLEQEQSDLEQDRLAKELK- 779
Qy 832 SIESKD---TEQKELQDMFTED---EDDPGSHKGRKRGKQNGKFKETRQEQINCVTRE 886
Db 780 ERRAKEKLEQEQSDLEQDRLAKELQEQSDLEQERRAKE-----KLEQEQSDLEQ 831
Qy 887 LLTAD---EEEALEKQEHORKEKELLEKIQSALACTNIPPLGRDRMYRRYWPISPGLE 944
Db 832 RLAKELQEQSDLEQERRAKEKLEQEQSD-----DLEQDRLAKELK-----LQ 873
Qy 945 EDYSGLTEDMLLPSPSSQNNVQSDPQVSTKTGTEPLMSESTSNIDQG--PRDHSVQLPK 1002
Db 874 EQSDLEQERRAKEKLEQEQSDLEQERRAKEK-----LQEQSDLEQERLAKELQEQ 928
Qy 1003 PVHPNRCWCFYSSCEQLDOLIEALNSGRHRESALKETLLQEKSRICQAQLARFSEKTFPS 1062
Db 929 DLEQERR-----AKEKLEQEQSDLEQERRAKEKLEQEQSDL---EQERLAKELQEQ 977
Qy 1063 DKPQDPSKPTYSRGRSSNAYDPSQWCAEQLELRLDFLLDIEDRIYGTIGAIKVTRH 1122
Db 978 QSDLEQERLAKELQEQSDLEQERLAKELQEQSD-----LEQERLAKELQEQ- 1028
Qy 1123 IWRSALESGRY--ELLSEENKENGIIKTWNEDVEEMEIDEQTKVIIVKORLLGIKTE-TPS 1179


```
Db 1042 QIRPIMERIKSSLS-----LRDHLSUTAFYKNE-----EKLKLEEN 1079
Qy 1098 RDFL-----LDIEDRIYQTLGAIKV-TDRHIWRSALSGRYELLSE-----NK 1141
Db 1080 NEFTENELNVSDMDVDDK-NSGVKSEVDVQVDAEKEKREAVIDEKLEVIADLMKLDSSK 1138
Qy 1142 ENGIKIYNE-DVEMETEDQTKVIVKDRLLGIKITTETPS-TVSNASTPQSVSSVWHYLA 1199
Db 1139 TRNVLNRIQELDQDELLEOKKSIIISQRPGARILASERKTKISRGKNVKNQIEILT 1198
Qy 1200 MALFOIEGIERRFLKAPLDASDSGRSVKTVLDR--WRESLLSSAS 1243
Db 1199 -----DLVNRHFKAMEDVI-----AWKNVLANSINGSSLRKNAS 1233

RESULT 19
S28261
centromere protein E - human
N:Alternate names: centromere 312K protein; kinesin-related protein CENP-E
C:Species: Homo sapiens (man)
C>Date: 03-Mar-1994 #sequence_revision 03-Mar-1994 #text_change 09-Jul-2004
C:Accession: S28261
R:Yen, T.J.; Li, G.; Schaar, B.T.; Szilak, I.; Cleveland, D.W.
Nature 359, 536-539, 1992
A:Title: CENP-E is a putative kinetochore motor that accumulates just before mitosis.
A:Reference number: S28261; MUID:93024922; PMID:1406971
A:Accession: S28261
A:Molecule type: mRNA
A:Residues: 1-2663 <YEN>
A:Cross-references: UNIPROT:Q02224; EMBL:Z15005; NID:g299864; PIDN:CAA78727.1; PID:g299865
C:Genetics:
A:Gene: GDB:CENPE
A:Cross-references: GDB:361164; OMIM:117143
A:Map position: 4q24-4q25
C:Superfamily: centromere protein E; kinesin motor domain homology
C:Keywords: ATP; coiled coil; microtubule binding; mitosis; nucleotide binding; P-loop
F:7-335/Domain: kinesin motor domain homology <KMOT>
F:86-93/Region: nucleotide-binding motif A (P-loop)
F:486-2183/Domain: coiled coil #status predicted <COI>
F:92/Binding site: ATP (Lys) #status predicted

Query Match 3.4%; Score 293.5; DB 1; Length 2663;
Best Local Similarity 18.7%; Pred. No. 5.8e-05;
Matches 305; Conservative 287; Mismatches 592; Indels 449; Gaps 74;

Qy 212 LTYQEALESEKKAQNLSQSPPELIIPVLYITSL-THR-----SRHEICDDIFAVVDKRY 266
Db 1144 LNVOEEMSEMKINEIENLNKNELNKNELTLEHMETERLELAQKLNENYEEVKSITKPK 1203
Qy 267 FVEETVEIRNNGARLOCTILEV-----LPPSHQNGFANGHV-----NSVDGETI 311
Db 1204 VLKELQKSFETERDLRGYIREIATQLOTKEELKIAHILKHEQETIDELRRSVSEKTA 1263
Qy 312 IISDSDSETOSCFSQN-----GKKQDAIDPLLPYKVQPTKKEL-----HESAIKATQI 362
Db 1264 QIINTQDLKSHTKLOEIPVLHBEQELLPNVVKVSETQETMNELELLTEQSTTKDSTTL 1323
Qy 363 SR-----RKHLFRDKLKLFLQ--HCEPOGVKIKKASSLSY-----KIAEQD 405
Db 1324 ARIMERLRLNEKFOEQSEETKSTUKERDNUKTKEALEVHDKQLKHEIRETLAKIQES- 1382
Qy 406 FSYFFPDPPPTFFIPANRRRGRPPKRIHISQEDNVANKQPLASVRSKATKERDKLKQE 465
Db 1383 -----QSKQEQLNMEKKNETK--IVS-EMEOPKPKDSALLRI 1419
Qy 466 EMKSLAFEKAKLREKADALBAKKKEKEDKKEKE-----ELKXIVBERLKKKEKE 518
Db 1420 EIEMLGL--SKRLQESHDEMKSVAKEDDLQLQLQEVLSQSDQLKENIKETIVAKHLETEE 1477
Qy 519 RLKVER-----EKERE-----KLRE--EKKKVEY 541
Db 1478 ELKVAHCCLKEQETINELRVLNSEKETEISTIQKLEAINDKLQNKIQIYEKEEQLN- 1536
```

```
Qy 542 LKQNSKPREDMECDLKLPEPTPVKTRLPPEIFGDALMVLE--FLNAGBELFDLQDE-- 597
Db 1537 IKQISEVQENY--NELQKFKEHRKAK-----DSALQSIESKMLELTLNRLQESQEBIQ 1586
Qy 598 --PFDGVTLVLEAL-----VGNSEGPLCELLFFLFTAIQATAEE--EE 640
Db 1587 IMIKEEEMKRVQBALQIERDQLKENTKEIVAKMKESQEKYQFLKMTAVNTEQMKCEI 1646
Qy 641 EVAKEQLTDADTKGCSLKLSDLSCTLSEILRLHLASGADVTSANAKYRVQKRGFPAT 700
Db 1647 EHLKEQF---ETQKLNLENIETENIRLTQILHEN-LEEMRSVT-----KERDRLSV 1694
Qy 701 DDACVELRLSNPSLVKKLSSTSVYDLTPGEMKILH-----ALCGKLLTLVSTR---- 749
Db 1695 EET---LKVERDQKLENLRETITRDLEKQELKIVHMLKHEQETIDKLRGIVSEKTNBI 1751
Qy 750 ----DFIEDYDILR---QAKQEFRELKABQHKEBEAAARIR--KRKEEKLKQEQ 798
Db 1752 SNMQKDLSEHNDALKAQDLKIQEELR--IAHMLKBEQOETIDKLRGIVSEKTDKLSNMOK 1809
Qy 799 -----KMKKEKQELKEDORNSTADISIGEER-----EDFOTSIESKOTE 839
Db 1810 DLENSNAKLQEKIQELKANEHQLITLKDVNETOKKVMSEMQLKQIKQDLSLTLSKLEIE 1869
Qy 840 QKELDDQMTDEDDPGSHKRRGRKGQNGFK-----EFTROEQINCVTRELLT 889
Db 1870 NLNLAQELHENLEEMKVMKERNLRRVEETLKLERDQKESLQETKARDLE-IQELKLT 1928
Qy 890 ADEEALQKHQKKEKELLEKIQSAIACTNIFPLGRDRMYRYWIFPSIPGLFIEDYSYG 949
Db 1929 A---RMLSKEH---KETVDKREKISEKT-----QISD 1956
Qy 950 LTEDMLPRSSFONNVQSDPOVSTKTGPELMSESTSNIDQGRDHSVQLPKPVKPNR 1009
Db 1957 IQKDL-----DKSKDELQKIQELQKQELQLLRVKEDVNMHKKINEMEQLKQK-QFENY 2010
Qy 1010 WCFYSSCE-----QLDQLIEALNSRGHRESAL---KETLQEKSRICQALAR-FSEE 1057
Db 2011 LC---KCEMDFNQLTKKLESLEEIRIVAKERDELRIKESLKMEROQFIATREMIARD 2067
Qy 1058 KPHFSDKPPQDSKPTYSRGRSSNAYDFSQMAE-KQLELR-----LRDFLLDIED 1106
Db 2068 RONHOVKE---KLLSDGQOHLMESLREKCSRIKELLKRYSEMDDHYECLNRLSLDLEK 2124
Qy 1107 RI-YQGTGLGAIKVTDRHIWRSALSS-----GRYEL--LSEENKENG1---IKTVNEDVE-- 1154
Db 2125 EIEFHRIMKKLVLSVTVKIKEQHECINKFEMDFIDEVEKQKELLIKIQLHQDDCDVP 2184
Qy 1155 -----EMEIDEQTKVIVKDRLLGIKITTETPSTVSTNASTPQSVSSVHYLAMA 1201
Db 2185 SRELRLDLKLNQMDLHIEILKOPSESEFFSIKTEFQOVLNKRKEMTQFLSEWLN---T 2240
Qy 1202 LFOIEQ---GIER-----RFLKAPLADSDSGRSYKTVLDRWRRESLLSSASL 1244
Db 2241 RFDIEKLNGIKENDRICQVNNPFNNRIIATINESTEFEBERSATISKEMEQDLKSLKEK 2300
Qy 1245 SQ-VFLHLSTLDRSVIWSKILNARCKIRCKKGDAENWVLDCGDRGHHTYCVRPKLTIV 1303
Db 2301 NEKLFFKNYQTLKTKSL-----ASGAQVNPFTQD--NKPHTVTSRATQLT-- 2342
Qy 1304 PEGDWFCPECRPKQRCRRLS-----FRQRPSSLESDEDEDSDSMGGEDE 1346
Db 2343 -----EKIRELENSLHEAKESAMHKESKIIKMQKELEVNDIIAKLQAKVHE 2389
Qy 1347 VGDDEEGQSEEBEYVEQDE-----DDSQEEEVSLPKRGRQPVRLPVTRGKLSS 1398
Db 2390 SNKCLEK---TKETIQVLQKVALGAKPYKEEIEDLKMVLVKIDLEKMKNAKSEKESISA 2446
Qy 1399 SFSRGGQOQEPGRY----PSRSQOSTPTTVTSSKTGSLRKINSAPT-----ETK 1445
Db 2447 TKATVEYQKEVIRLLRENLRSSQAQDTSVISEHTD---PQPSNKPCTCGGGSGIVQNTK 2503
```

```
Qy 1446 SLRIASRSTRSHGFLQADV-----ELLSPPR-----KRRG 1478
Db 2504 ALILKSEHIR-----LEKEISKLQKQNEQLIKQKNELLSNNHLSNEVKTWKERTLKREA 2558
Qy 1479 RKSANNTPEPNPFRVATKSSBGRSNIASKLSLQ-----SESKRRC-----RKQS 1531
Db 2559 HKQV--TCENSPKSP--KVTGTASKKQITPSQCKERNLQDPVPKESPKSCFFDSRSKSL 2614
Qy 1532 PEPSPVTLGRSS 1544
Db 2615 PSPHPVRYFDNSS 2627

RESULT 20
T18296
myosin heavy chain - Entamoeba histolytica
C;Species: Entamoeba histolytica
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C;Accession: T18296
R;Guillen, N.
submitted to the EMBL Data Library, February 1997
A;Reference number: Z18865
A;Accession: T18296
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-2139 <GUI>
A;Cross-references: UNIPROT:Q07569; EMBL:I03534; NID:g1850912; PID:g1850913; PIDN:AAB480
C;Genetics:
A;Gene: mhca
A;Superfamily: myosin heavy chain; myosin motor domain homology
F;91-780/Domain: myosin motor domain homology <MWO>

Query Match 3.4%; Score 289.5; DB 2; Length 2139;
Best Local Similarity 18.4%; Pred. No. 6.7e-05;
Matches 274; Conservative 237; Mismatches 540; Indels 437; Gaps 57;

Qy 152 PLLHRKPFVQKPPADLRPPBEEFYKVTNEIPRHYDDFFERTILCNLSVWCAVTPRG 211
Db 852 PMLKRRPFEGKKD--KEIEDLKKLAEIEKK-----REAAENALASATAKTELE 902
Qy 212 LTVQE-----ALESEKKA-----RQNIQSPPELIIPVLYLTSITH-RSRLHEICDDIFA 260
Db 903 AKIQDLEDKISELSKLSAAELDKQELN-----LKIENLEEDKEELKETIDNLKG 952
Qy 261 YKDRYFVEETVEIRNNGARLQCTILEVLPPSHQNGFANGVNSVD----- 307
Db 953 DLKDSKLKGEDLEV-----EITELNSQINTLNATVDDKDKTIAEMQESIDEK 999
Qy 308 -----GE-TIISDSDDSE-----TQSC 324
Db 1000 EDEITKLGIKLLSEKDDLEQDRADVSATKDIAKLNKIITECEDAKDEAKLEQEL 1059
Qy 325 SFQNGKKDAIDPLLFYKVQPTKKELHESAIVKATQI-SRRKHLFSRDKLKLFLKQHC 383
Db 1060 EDEENKNKDLTNEL-----QOTQLKGTEKSLAAQVAATKASDERDTSQNLNEKL 1113
Qy 384 PQEVIKIKASSLTYKIAQDFSYFFPDPTFIPSPANRRG--RPPKRIHSQDNV 441
Db 1114 TTKNLTKTKADLEKKISGLQDYEDLEDD-----KNKIEGDLRNAQRKIKELDDEI 1164
Qy 442 ANKQTLASYSKATPKERDKLL--KQEMKSLAFE-----KAKLKREK-----ADALEAKK 489
Db 1165 TKGADVSOYLQKQEEVESQIAKKQEEKEAIGNDVKNKEKTIKEKELEISLQEKLDTE 1224
Qy 490 KEKEDKEKREELKKIIVEERLKKKERLKVREKERELKEERKRYVEYLKQWSKPR 549
Db 1225 VEREDAEEKKKE-----IEKEMKALQEEKENV-----ESSKNSTEDKKKKLEDNLKDTQKKL 1276
Qy 550 EDMEDCDLKEPPTPKVTRLPPIFGDALMVLEFNAGELFDLQDFPDGVT-LEVLE 608
Db 1277 DDMTADN-----EKLKAKAK-----DLEAQLNEVQDNHKEKAVADAELLN 1315
Qy 609 EALVGNDSGLPCLELFFFLTAIFQAI-----AEEBEEVAKELQTDADTKGCSLKSLDL 662
```

```
Db 1316 KKQAQSKELNSLKAELEALTAKSVVESKNKQSENEKAALSEBIDQANEK---LKNIOA 1372
Qy 663 DSCITLSEILHLILASGADVTSANAKRYQKRGFDATDDACMELRLSNLSLVKKLSTSS 722
Db 1373 D-----LRKATADLQEAN-----EKKAEEVEAQRD---KLVDANKMTKMTLEIK 1413
Qy 723 VYDLTPGEKMKILHALCGKLLTLVSTDFIEDYVILRQAKQEPRELKAEQHRKEREEAA 782
Db 1414 ARDEENTYK-----VENTEKVLKRKEADLEE-----ANENLDIE 1447
Qy 783 ARIRKRKEEKLKEOEQKWKKEQEKL-----KEDE-----QRNSTADISIGBEEREDFDTSI 833
Db 1448 KQORMNKEKQVKLEGELKETKQKLNAAIAEKDSIFTAKKQSDADLELNKTVTEHDEVV 1507
Qy 834 ESKDTEQEKELDQDMFTEDDDPSGHRKGRKGQNGKFEKFTROEQIN-CVTRELLT--A 890
Db 1508 AKLNTQITKLTRDNQSAEEE-----LNELSKADKDKKKISELEEQVNELESFVGTGNA 1562
Qy 891 DEEE-----ALQEHQORKEKELEKLEKIOSAIACINIFPLG 924
Db 1563 DENEIKIRDAQIADLNKALEMKGVQNNQLQATNKELKAKONDLSKIE----- 1610
Qy 925 RDRMYRYWIFPSIPGLFIEEDYSGLTEDMLPRPSSFQNNVQSQD-----POVSTK 976
Db 1611 -----ITENEMKKLENAKKLEQDKDEADKAVSQTIKRRK 1645
Qy 977 TGEPLMSESTNIDQGRDHSVQLPKPVHKNWCFYSSCEQLDQLEALNSRGHRRESAL 1036
Db 1646 GLBEEVKKLTTETQ-----ALKFOINAPSSVAQEEKQRLSEADIAELKEQLEQERTT 1697
Qy 1037 KETLLQEKSRICAQLARFSEKHFSDKPOPDSKPTVSRGRSSNAYDPSCWCAEQLELR 1096
Db 1698 AANAERKKIQAEI---DEVKNLEEDVTNQREKLV-----AKNSEDAEISLUKEKKA 1749
Qy 1097 LRDFLDIEDIRYQGTGALIKVTDRIHWRSALESGRYELLSEENKENGIIKTVNEDEVEM 1156
Db 1750 LED---EIE-----KITDDNNKLESE-----EIDSLORKYNALLDSKDSVSMK 1789
Qy 1157 EIDEQTKVIVKRLLGIKTETPTSTVNSTAPQSVSVVHYLAWALPQIE-QGIERFLK 1215
Db 1790 EKQDELKVTKDALETEKKNHAETMRLKGRLEKAAEV-----QVRLEALQKNL 1840
Qy 1216 APLDASDSGRSY-----KTVLDWRRESLSSASLSQVFLHLSLTD--- 1255
Db 1841 AQOEKAKATKYRAADGELSKLMNELDDVKDQDLKADQDLADKED-----ELATLQKY 1894
Qy 1256 RSVIWSKSIILNARCKICRKKGDAENMVLCDCGRGHHTYCVRPKLKTVPEGDMFCPCRP 1315
Db 1895 KTLVKQKSVFDSRIQEMQEQDLLEK-----AGRA 1923
Qy 1316 KQRCRLSPQR--PSLESEDEVED-----SMGGEDEVGDDEEGQSEBEEY 1361
Db 1924 KAQKQKQAYEKKLQELQENDNDNDFEYKETADKRIINTLSAQKDLQKEKELGKQSEK 1983
Qy 1362 EVQDEDDDSQEESEVSLPKEGRPQVRLPVKTRGKLSSSFSSRGQQQPGYPSPSQSTP 1421
Db 1984 EVORLRVKCO-ELETKVAVGGANVSI-AKVAKYAEIEBELTTEADALKAKMKAEEKKA 2041
Qy 1422 KTVTSSTKGRSLRKINSAPPT---ETKSLIRASRSTRSHGFLQADV 1465
Db 2042 KT-----SOKKLDELQKTIADYETKE---ASFNT--EIGKTQAL 2076
```

RESULT 21

S67593
transport protein US01 - yeast (Saccharomyces cerevisiae)
N;Alternate names: protein D2552; protein YDL058w
C;Species: Saccharomyces cerevisiae
C;Date: 12-Jul-1996 #sequence_revision 12-Jul-1996 #text_change 09-Jul-2004
C;Accession: S67593; A38455; S30782
R;Bloecker, H.; Brandt, P.
submitted to the Protein Sequence Database, July 1996


```

Query Match      3.3%; Score 285.5; DB 2; Length 1871;
Best Local Similarity 18.7%; Pred. No. 8.6e-05;
Matches 365; Conservative 306; Mismatches 705; Indels 577; Gaps 90;

Qy      1  MEDASESRGVAPLINNWLPGPSPLSPVSVGTCKSHRVANKKVEARSEKLLPTALPPSE 60
Db      101  IDERIEEKGGLAD--SNKESVDSLSRPPDI EGRECHEQTTHEEQENNKQLVQAESDSD 158
Qy      61  PKVDQKUPRSESR-----RGSGGGTOFFPARSRAVAAGEAAARGAAGPGRSGPLGRVSP 114
Db      159  DFGSRAPEETEEQESDVLDRSTSTSGAMEKEMTDVGDGLRKVQGEIEPEPHNE--ESKISE 217
Qy      115  RCLCSGB--GGQAVGVGIAGRGGRGDSRRAPGGREMPLLHRKPPVRQKPADLRPDE 172
Db      218  --MVDGTSHEKKKVVMDKORDVKEEDVGAMGEGFRNIDRTQVVGDDETAETEKND 275
Qy      173  EVFYC-----KV-----TNEIFRHYDDFFERTILCNLSLVMSCAVTGRPGTYQEALE 219
Db      276  EEFESDKLEADVDKINEGGNTKVRHSED-----ENLKLQEKKEE 316
Qy      220  SEKKARQNLOSPEPLIIPVLYLTSLTHRSRLHEICDDIFAYVKDRYFVEETVEVIRNN-- 278
Db      317  QHSKEQKG-----HSKEENMKELVEEK--TPEAETTRINDI 350
Qy      279  ---GARIQCCTILEVLPSSH-----QNGFANGHVMSVDGETIIISDSDSET 321
Db      351  LGPQGEIEVPEVDTLGKTSDEGKEKQNI VKKEIKNGDATSEIDAKWGE--VPASNIADTGM 409
Qy      322  QSCSFQNGK--KXDAIDPLFKYKVQTKBELHESAIVKATQISRRKHLFSRDKLKLFLK 379
Db      410  NSEDFESDKLESADENVDMVEK---KDRQENDKVAQSEDISLTQLQBIGEQQFOGQK 465
Qy      380  QHCEPQGVGIKASSLSTVKIABODFSYFPDDPPTFIFSPANRRGRPPKRIHISOED 439
Db      466  RH--DKQENIKELREGQASE--AEKNIK-----NDILKPQKRGSEGHKTKQTFQEE 513
Qy      440  NVANKQTLASYSRK-----ATKERDKLLQBEEMKSLAPEKAKLRKEXADALEAK 488
Db      514  --TNKQP--EGYNEKIMETGKKNIBDGTGRKVQEMIRQOELD-----EPARSEKENR--SREL 565
Qy      489  KKEKEDKKEKEELKKTIVBERLKKKEERLKVVEREKE--REKLREKRKKVVEY--LKQ-- 544
Db      566  KSKTNDSEKKE---KETAGTER--KEKESDRPKLTREQEVAEDAEKDTKFSYGEVKEE 620
Qy      545  ---WSKPREDMECDDLKELPEPTPVKTRLPPEIFGDALMVLEFLNAPGELFDLQDEPPDG 601
Db      621  BEIAGKEKEFGSDDDIARVDT-----EQLDSNAMKGQEEKMDIQELV--LEEKYCDG 672
Qy      602  -----VTILEVEEALVGNDSBGLPCCELLFFPLTATFOAIAE----- 637
Db      673  KGKIIAETAETKAENNKSRVQETEEOQLKDREDTCG-----KHFOKLIEGETISDH 721
Qy      638  -EEEV-----AKEOLTADTKGCSLKSLDL-----DSCITLSEILRLHLIASG-- 679
Db      722  GEVEDVEKGRKTEAEKRIKDRABAEIEIKEDLVSGRYIKGTTIKELVENRGIRYNEH 781
Qy      680  -----ADVTSANAKYRQKRGFPDAT-----DDACMEL 707
Db      782  EEKKKDDANRPKITGTIKQELVSLNSQLQENVEDGDKTQELVEEKIKDCBEEGSEES 841

```


Qy	531	L-----REKRKYVEYLKQWSKPREMECD-----DLKELPEFPVKTRLPPEIFGD	577
Db	200	LIELEGLKRSABEAQKFEELHKQ--SASHADSESOKALEFSSELLKSTKESAKEMEEXKAS	258
Qy	578	ALMWLEFLN-----AFGELFDLODEFDPDGTVLEVLLEALVGNDSSEGPLCE	622
Db	259	LQOEIKELNEKMSNEKVEAALKSAGSALAABQEE-----LALSKRLLTETQOKVSSTE	312
Qy	623	LLFPFLT-AIFQAIABEE---BEEVAKBQLTDADTKGCS-----	656
Db	313	ALIDELAQLEQKKASERSRFEELSVLQDLDAQTKGLQAKLSEQEGINSKLAELKEKEL	372
Qy	657	LKSLDLD-----SCTLSEILRLH--ILASGADVTS-----ANAKRYQKRG--G	696
Db	373	LESLSQOEKELRTANESKLAEBVLKEKEALEANVAEVSNTVATVEVCNELEELKLTSDEN	432
Qy	697	FDATDDACMELRLNPSILVKKLSTSVVDLTPGKMKILHALCGKLLTLYSTRDF-I	755
Db	433	FSKTDALLSQALSNNSELEQKLK-----LEELHSEAGSAAAAATQKNLEUDV	481
Qy	756	V-----DILQAQKQBFRELK-----AEQHRKEREBAARIRKR---BEKLEQEQKMKK	803
Db	482	VRSSQAABEAKSQIKELETKFTAEOKNAELEQQLNLQLKSSDAERELKELSEKSEL	541
Qy	804	QEKLK--BDEORNSTADISTEEREPDFTSIESKOTEQEKELDOOM-----FTEDDDP	855
Db	542	QTAIEVAEAEKKQATTOQEQYKOKASELLESLTOSTSARNSELEEDRLALQKGAHEHRA	601
Qy	856	G-SHKR-----GR-----	862
Db	602	NTTHQRIELEGLOQSSQSHEDAEGRKLKOLELLQTEKYRIQBELEQVSSLEKKHGETE	661
Qy	863	---RGKRQ-----NGKFEPTQEOINCVTRELLTAADEE	893
Db	662	ADSGYLGQVAELQSTLEAFQVKKSSLEAALNIATENEKELT--ENLNAVTS-----KKKL	716
Qy	894	EALKQEHQKKE---LLEKIQSAIACNTFPLGRDRMYRYWFPPIGFLFEEDY--S	948
Db	717	EATVDEYSVKISENLESLIRNELANTQO---CKLES-----IENDLKAA	758
Qy	949	GLTEDMLPRPSSPQNNVQSDPOVSTKTGEPMLSES---TSNIDQGRDHSVOLPKPVH	1005
Db	759	GLQSEVMEKLSAESLEQKGRIDEATYKMELEALHQSLSDS---EHRLQ--KAME	813
Qy	1006	KPNRWCPYSCQDLQDILALN-----SRGHRESALKETLLOEKSRICAQL	1051
Db	814	E-----FTSRDSEASSLTEKRLDLEGKIKSYEOLAEASGSSSLKEKLEQTLGRLLA--	866
Qy	1052	ARFSEKHFSDKPPQDSKPTYSGRSSNAYDPSQMACB--KQLELRDLRDLDDIEDRIY	1109
Db	867	ABSVNEKL-----KQEFQO---AQEKSLQSSSESELLAETNNQIKIOEL-----	909
Qy	1110	QGLTGAIKVTDRIHWRSALESGRVYELLENKENGII---KTVNEDVEEME-----I	1158
Db	910	EGLIGSSV-EKETALKRLEAETRFNQKETESSDIVEKLKTHENOIEEYKLAHEASGV	968
Qy	1159	DEQTKVVKORLLGIKGTETSTVSTNASTPQSVSSVVHYLA-----MALFOIEQGIERRF	1213
Db	969	ADTRKVELEDALESLKLNLESTIEELCAKCGLEKESGDIAEVLNKLNLALNHSSEANE	1027
Qy	1214	LKAPLDASDSGR-----SYKTVLDRWRSELSASLQVFLHLSLTD--RSVIMSKS	1263
Db	1028	LQTKLSALEAEQOTANELAEASKTTIIBDLTKQLTSGEKUIQO-----SOIEKLRAVAEKS	1082
Qy	1264	ILNARCKICRK-----KGDANNMVLCDGCDRGHHTYCVRPKLTVEGDWFCPECR	1314
Db	1083	VLESHFELEKLTSEVKAQLKENVENAA-----TASVK-----VAE-----	1118
Qy	1315	PKQCRKLSFRQPSLSEDEVDMSMGEDDEVDGDEEGOSBEEVEVEQDEDDSDQEE	1374
Db	1119	-----LTSKLOEHHIAGERDVLN---EQVLQLOKQAAQSSID--EQK	1158

Qy 1375 EVSLPKR-----GRPVRLPVKTRG-----KLSSSFSSRGQQBPFG 1410
Db 1159 RLKLRKLSLNLNQSKILARKCSQSMQLKRLTEAMDVGKSRDLIDLSFSSP----- 1211

Qy 1411 RYPSRSQSPTKTIVSSKTGRSLKINSAPPTETKS 1446
Db 1212 --TKRKSKRPASLSSSSG---NVTPTQTAS 1241

RESULT 26
S28589
trichohyalin - rabbit
C:Species: Oryctolagus cuniculus (domestic rabbit)
C>Date: 12-Mar-1993 #sequence_revision 01-Mar-1996 #text_change 09-Jul-2004
C:Accession: S28589
R:Fietz, M.J.; Rogers, G.E.
Submitted to the EMBL Data Library, December 1992
A:Description: Examination of the gene encoding rabbit trichohyalin.
A:Reference number: S28589
A:Accession: S28589
A:Molecule type: DNA
A:Residues: 1-1407 <FE>
A:Cross-references: UNIPROT:P37709; EMBL:Z19092; NID:g1746; PIDN:CAA79519.1; PID:g1747
C:Comment: Trichohyalin is a protein of the medulla of the hair and of the inner root sheath. Covalent modifications to this protein include conversion of arginine to citrulline and acetylation.

C:Genetics:
A:Introns: 46/3
C:Superfamily: trichohyalin; calmodulin repeat homology
C:Keywords: calcium binding; citrulline; EF hand; hair; tandem repeat
F:49-81/Domain: calmodulin repeat homology <EF2>

Query Match 3.2%; Score 280; DB 1; Length 1407;
Best Local Similarity 18.5%; Pred. No. 0.00011;
Matches 245; Conservative 218; Mismatches 440; Indels 422; Gaps 50;

Qy 345 OPTKKELHE--SAIVKATQISRRKHFLFSRDKLKLFLOKHCPEQGVIKIKASSLSITYKIA 402
Db 331 QOLKRELEIRERQRLEQEERBQLAAEE---VREQARERGESLTRWRQQLSEAGA 386

Qy 403 EQDSFYFPDDPPFIIFSPANRRGRPPKRTHI-----SQEDNVANKOTLAS-- 449
Db 387 RQSKVYSRPRRQEQSLRQDQERERRELBEEAQRRQQWQAEEEEERRRQSLARP 446

Qy 450 -----YRSKATKERDKLLKQE-----EMKSLAPEKAKLKREKADALEAKKKKEKD 494
Db 447 SLRERQLRAEERQEQEQFRFEEERRRERRELQFLFEEEQLORRERAQQIQEEDSFQED 506

Qy 495 KEKKREBLK-----KIVEERLKXKXEKPERLVKVEREKREK 530
Db 507 RERRRRQEQRPQGTWRWLQEEAQRRHTLYAKPGQEQQLREBEELQREKRRQREREY 566

Qy 531 LREKRVKYVEYLKOWSPREDMECDLDELPEPTVPKTRLPPEIFGDALMWLEFLNAFGE 590
Db 567 REEEKLOREDEKRRQERE-RQYRELEEFLRQEBRLDR----- 604

Qy 591 LFIDLQDFPDGVTLVELLEALVGNDSGLCELLFFFLTAFQAIABEEVEAKE---QL 647
Db 605 --KLREE-FOLLQEREERLRQGERKLER-----FEQLLRQEBELQREKRL 652

Qy 648 TDADTKGCSKLSDSCITLSEILRLHLIASGADVTSANAKYRYQKRGGFDPATDDACHMEL 707
Db 653 REEE-----QLLR-----REEQLRQERERKLREEEQQLQER 684

Qy 708 ---RLSNPSLVKLLSSITSVDLTGCRKMILHALCGKLLTLVSTDFDYDVILLROAKQ 764
Db 685 EEERLRQRERAKL----- 710

Qy 765 EFRELKAEOHKKEREE-----AAAIRKKRKEELKEQEQWKKEKOE----KLKDEORN 814
Db 711 ELRO---ERKKUREEQQLRREQLLRQERDRKLREEEQLQLESEEERLRRQEREQQLR 767

Qy 815 STADISICEEP---REDFDTSIESKQTEQKELDODMTFEDDDPGSHKRGGRGK-RGQN 870

Db 768 RERDRKFEQQQLQEREEERLQRERKRLREBEQQLQEREBERLQRERKRLREBEQ 827
Qy 871 FKEFTROQINCVTRELLTADDEEALKQEHQ-----RKEKELLEKIOSAIACINI 920
Db 828 LLQREERLRRQERKRLREBEQQLQEREBERLQRERKRLREBEQQLQEREBEQLR 883
Qy 921 FPLGRDRMYRWIIFPSIPGLFIBEDYSGLTEDMLPRPSSFQNNVQSDPQVSTKTGEP 980
Db 884 --QERDRKLE-----EEQLRQEE--QELRQERDKL--REBEQ 917
Qy 981 LMSFS-TSNIQDGRDHSVQLPKPVHKPNRWCFYSSCEQLDQLEALNSRGRHSALK-- 1037
Db 918 LLQSEERLRRQERKRLR-----EEQQLRREBEQELRRERARKLR 959
Qy 1038 --ETLLOE-----KSRICQAQRFSEKTHFSKDPQDPSKPTYSRGSSNAY 1082
Db 960 EEQLQEREBERLRRQERARKLREBEQLLRREBEQELR-----QERDRKFR 1005
Qy 1083 DPSQWCAEKQLELRDLFLDIEDRIYQGTLGAIKVTDRIHWRSALESGRYELLSEENKE 1142
Db 1006 EEEQLQERE--ERLR--RQERDKF-----REERQLRQEELE--QFQERDRK 1051
Qy 1143 NGIITWNEDVEIMEIDEQTVIKVDRLLGKTTETPTVSTVSTASTPOSVSVVHILAMAL 1202
Db 1052 FRLEEQIRQEEKQLRQEE--RDR-----K 1075
Qy 1203 FOIEQGIERRFLKAPLDASDSGRYKTVLDPRWRESLLSSASLSQVFLHLSTLDRSVWSK 1262
Db 1076 FREEQORRRQERQELRRERDRKF-----REBEQLQEREEER--LRQERARKLREE 1128
Qy 1263 SILNARCKICRKGKD---AENMVLCDGCD---RGHHTYCVRPKJKTVPBGDWFCPECR 1314
Db 1129 QLLRREBQLQERDRKFRREBEQQLQEREBERLQRERKRLREBEQQLQEREB--ERLR 1186
Qy 1315 PKQCRRLSFQRPSLESDEV-----EDSMGGEDEVGD-----EEGQSE 1357
Db 1187 RQERARKLREBEQQLQEREBERLQRERARKLREBEQQLQEREBERLQRERDRKFRREEQLLR 1246
Qy 1358 EEEVEVEDQSDQSEBEVSLPKGRQVRLPVKTRGKL-----SSFSRGGQOE 1408
Db 1247 REQELRRERDRKFRREBEQQLQEREBERLQRERAR--KLREBEQQLFEEQEEQLRQER 1305
Qy 1409 PGRYPSQSQTPTKTVSSKTGRSLKINSAPPTETKSLRIASRSTRSHGPLQADVFE 1468
Db 1306 DRRY--RAEQFAREERKRLERELRQ-----E 1331
Qy 1469 LLSPRKRGRKRSANNTPENSPNPNRVIATKSSEQSRSVNIASKLSLQSESKRCKR 1528
Db 1332 EQRRRRERERK-----FR-----EQELRR-----QOEERQRRQLRE 1364
Qy 1529 RQSPF 1533
Db 1365 RQFRE 1369

RESULT 27
T21809
hypothetical protein F35G12.8 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T21809
R/Chui, C.
submitted to the EMBL Data Library, October 1994
A:Reference number: Z19473
A:Accession: T21809
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1549 <N1>
A:Cross-references: UNIPROT:Q20060; EMBL:Z46242; PIDN:CAA86336.1; GSPDB:GN00021; CESP:F35G12.8
A:Experimental source: clone F35G12
C:Genetics:
A:Gene: CESP:F35G12.8
A:Map position: 3

A:Introns: 50/3; 72/3; 177/3; 235/3; 333/2; 581/3; 667/3; 812/3; 882/1; 1050/3; 1223/3;
C:Superfamily: chromosome segregation protein SMC1

Query Match 3.2%; Score 280; DB 2; Length 1549;
Best Local Similarity 18.1%; Pred. No. 0.00012;
Matches 293; Conservative 253; Mismatches 594; Indels 480; Gaps 65;

Qy 153 LLHRKPFVRQPPADLRPDDEVFYCKVTNBIFR-----HYD-----DF-PERTILCNLSL 200
Db 153 LIHKS-----AGRPDK-----CTVIHFQRLVDIPGHYEVVKDSEFTISRATFQNN 200
Qy 201 VMSCAVTRPGLTVQEALESEKARQN-----LQSPPEPLII--PVLVLTSL 245
Db 201 S-SVAIDGRP-----ATKNEVEARLRVDIDIEHNRFLILQGEVEQIAMKPVKTKSE 253
Qy 246 THSRRLHEICDDIPAYVKDRYFVEETVEVRNNGARLQCTILEVLPSHONGFANGHV-- 303
Db 254 T---GMVEYLEDITGTNR-----LEPFVKLFQRRVNRILTCDL-----SQRIARDHARN 299
Qy 304 -----NSVDGETIIISDSDSETQSCSFQNGKKKDAIDPLLPKYKVQPTKKELHE---- 353
Db 300 SKVAMENPVRAAIEFLMKENEATTIHKLEQRQRVILD-----KIAPKAELDKMKEE 353
Qy 354 -SAIVKATQISRRKHLFSRDKLKLFLKHOCEPOGV--IKIKASSLSYTKIAEQDFSYFF 410
Db 354 MKSIAETLDTNKNBYKQSEBAQKVMIBERSKLDKNFDSLSELSKELSDLGTEETRRKE---- 408
Qy 411 PDDPTTIFPANRRRRPPKRIHISQEDNVANKQTILASYSKATKERDKLLK-QEEMKS 469
Db 409 -----ALKRHOANISKAEEKEKYGKSNLEAAPEKAER---KIAKQOEVEQ 454
Qy 470 LAPEKAKLKEKAD-----ALEAKKEKEDKEKREELKIVKEERLKKKEEK 517
Db 455 L-LBIKTANEADKNLDEFKSEAPKEOKKIQTWAQKSNFNVKRGEGARI-AREDF 512
Qy 518 ERLKVEREKERKLEERKRYVEYLKQMSKPREME-----CDDLKELPEPTP--- 565
Db 513 EDLKLANSCTDKLIELKRLSESESYAKEDKDKLPKPFDSWMDKJKQLSTELPTLR 572
Qy 566 -----VKTRLPPEIFG-----DALM-----VLEFLNAFELFDLQ 595
Db 573 NTAQKNQDLAKTRDLRLETLRQQNSSSSSNKVTQALMKEKEAGRIKSFHGLRDLGVID 632
Qy 596 DEFPDGVTVLEV---LEELAVGNDSEGPLCELLFPFLTAIPOAIEEVEEVAKEQLTADT 652
Db 633 PKYEGACTNFGARLNLYLVKKEED---AKNVINFLVA-----NKLPRQTVQPLDK 680
Qy 653 KGCSLKSL-----DLDSK-----TLSEILRLHLASG 679
Db 681 IKCDKRDLANPTNPLPAPRLIDLIDCDPVLKPAPYDMVRSALVGDSTQEAQRMHRMPAC 740
Qy 680 ADVTSANAK-YRYOKRGGFADTDACMELRLSNPSLVKLSSTSYDLTPQEKMKILHAL 738
Db 741 RGVTVCTLEGSMIHPSGSFTGGKTVKGLILTDKNMAK-----QVTPEDK----- 786
Qy 739 CGKLLTLVSTRDFIEDYDILRQAKQFRELKASQHRKEREEREAARIRKKEELKQEQ 798
Db 787 -----AERDLAEK-----LGKLRDEADELKGQEHMDGQLEIAR-----RKVAEMSN 829
Qy 799 KMKEQKQELKDEQORNSTADISIGEEREDPDTIESKDTQEKELDDQDMFTEDDDPGSH 858
Db 830 RLSIVTSSVQSAAPAIETLKTIANQEAQKVKVDAKTLEDK---QKIVELEK----- 881
Qy 859 KRGRRGKRGQNGFKFTRQEQINC-----VTRELTADAEEALK--QBHQKKEKELLEKIQ 912
Db 882 ---KRDELGEAAKVKARQAEIQSKLDGIFKELVQCHRDEAKESLQKQKLEKDIKAE-- 936
Qy 913 SAICTNIFPLGRDMRYRWIIFPSIGLFIEDYSGLTEDMLLPRSSFQNNVQSDPQ 972
Db 937 ---TANISGRN-----TAKDENISRHDKD 960
Qy 973 VS---TKTGEPLMSESTSNIDQGRDHSVQLPKPVHKPNRWCFYSSCEOLDOLIBALNSRG 1030

Db 961 TEKMKKKCEELMEKA---ID-----DEEVSKK-----ETVEFEKQIKLQTKG 1002
Qy 1031 HRESALKETLLOEKSRI CAQLARFSEKXPHFDKPPDPSKPTYSRGRSSNAYDPSQMAE 1090
Db 1003 EBMTKQSLSAETKLEGELKKCE-----GI 1030
Qy 1091 KQELRLRDFLLDIEDRIYQGTGLGAIKVTDRIHWSALESGRYELLSSENKENGIIKTVN 1150
Db 1031 KELKESMLADRLKVED--IEKKLAALKVNRIPRFQFLIESRPEDL--EMQIDDKMPYVD 1086
Qy 1151 EDVEEMEIDEOTK--VIVKORLLIKTETSTVSTNASTPOV-----1191
Db 1087 ENQSEEVERQKHWACVMSDAAYALEFEMQKVLNTESEVNDGEDRVPVELLSDRKI 1146
Qy 1192 ----SSVHYLAMALFOIEQGIERRFLKAPLADSDSGRSYKTVLDRMRESILSSASLSQV 1247
Db 1147 NEISRDABEEMQMLKVCEQQVEA--LKAQVDIS--SIRAYVDVKVQYNEQVILKTIATEV 1203
Qy 1248 F-LHLSTLDRSVINSKSLNARCKICRKGDAENNVLCDCGRGHHTYCVRPKLKTVPEG 1306
Db 1204 HRKHNQELQRIKQMRLEBFHSAFEGFHLVAVFKMLTDGGD-----AKLEYIDKD 1254
Qy 1307 DWFCECPKQRCRLSRFRQPSLESDESDVSDMGED-----1344
Db 1255 DPF-----RQGISFWAPPAKAWKQIQFLSGGEKTLSSALIFALHMFRTPTFVYM 1305
Qy 1345 DEVGDEEGOSEEBEYVEQDEDDSQ-----1371
Db 1306 DEIDAALDYRNVSIIAQVVRQKTENNAQFIILSRNNMFELANRLVIGYKVDGTRNVAID 1365
Qy 1372 -----EEVEVSLPKGRQVRLPVKTRGLKSSFSR-----GQOQEPGRVPSSRQSGTP 1421
Db 1366 PLRVCMAKQITDSLGQATCTLPDEVQTRFNETWSRQNKEMIAOEKQVNPFPSSNEISKA 1425
Qy 1422 KTVSSKTRGSLRKINSAPPETKSLRIASSTRSHSGHPLQADVVELLSPRRKRGRKS 1481
Db 1426 EKIWNVE--GRVRKEL--IQTRDVTSRQSKATSGDG-----TERPASRS 1468
Qy 1482 ANNTPE-----NSGPNFNFRIATKSSQSRSVNIAASKLSQF--SESKRRCKRQSPSPSPV 1537
Db 1469 ASR--PESRINQMYPAPRLVRSQQNVSRPKARNIEADETTPPSKESNASITPKRSPM 1527

RESULT 28
A61231
N:Altitute: human
N:Alternate names: cellular myosin heavy chain; myosin type 9; NMHC-A
N:Contains: myosin ATPase (EC 3.6.4.1)
C:Species: Homo sapiens (man)
C:Date: 12-May-1994 #sequence_revision 14-Jul-1994 #text_change 09-Jul-2004
R:Accession: A61231; A34876; I52562; I61692
R:Simons, M.; Wang, M.; McBride, O.W.; Kawamoto, S.; Yamakawa, K.; Gdula, D.; Adelstein, C.; Res. 69, 530-539, 1991
A:Title: Human nonmuscle myosin heavy chains are encoded by two genes located on different chromosomes
A:Reference number: A61231; MUID:91316803; PMID:1860190
A:Accession: A61231
A:Molecule type: mRNA
A:Residues: 1-715 <SIM>
A:Cross-references: UNIPROT:P35579; UNIPROT:Q9UMJ0; GB:M69180; NID:g189029; PIDN:AAA6176
R:Saez, C.G.; Myers, J.C.; Shows, T.B.; Leinwand, L.A.
Proc. Natl. Acad. Sci. U.S.A. 87, 1164-1168, 1990
A:Title: Human nonmuscle myosin heavy chain mRNA: generation of diversity through alternative splicing
A:Reference number: A34876; MUID:90138958; PMID:1967836
A:Accession: A34876
A:Molecule type: mRNA
A:Residues: 715-1361 <SAE>
A:Cross-references: GB:M31013; NID:g189035; PIDN:AAA6349.1; PID:g189036
R:Toothaker, L.E.; Gonzalez, D.A.; Tung, N.; Lemons, R.S.; Le Beau, M.M.; Arnaout, M.A.; Blood 78, 1826-1833, 1991
A:Title: Cellular myosin heavy chain in human leukocytes: isolation of 5' cDNA clones, cDNA sequencing, and expression analysis
A:Reference number: I52562; MUID:92003925; PMID:1912569
A:Accession: I52562
A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA
A:Residues: 1-52, 'EAI', 56-659, 'T', 661-868, 'T', 870-930, 'C', 932-1239, 'KG', 1242-1337 <RES>
A:Cross-references: GB:M61105; NID:g188988; PIDN:AAA59888.1; PID:g553596
R:Bement, W.M.; Hasson, T.; With, J.A.; Cheney, R.E.; Mooseker, M.S.
Proc. Natl. Acad. Sci. U.S.A. 91, 6549-6553, 1994
A:Title: Identification and overlapping expression of multiple unconventional myosin genes in Dictyostelium
A:Reference number: A55758; MUID:94294418; PMID:9022818
A:Accession: I61692
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 182-218 <DEM>
A:Cross-references: GB:L29141; NID:g457249; PIDN:AAA20904.1; PID:g531134
C:Genetics: C
A:Gene: GDB:MYH9
A:Cross-references: GDB:120216; OMIM:160775
A:Map position: 22q12.3-22q13.1
C:Superfamily: myosin heavy chain; myosin motor domain homology
C:Keywords: actin binding; ATP; coiled coil; hydrolase; methylated amino acid; nucleotide binding site
F:184-764/Domain: myosin motor domain homology <MMOT>
F:174-181/Region: nucleotide-binding motif A (P-loop)
F:552-565/Region: actin binding #status predicted
F:626-640/Region: actin binding #status predicted
F:837-1938/Domain: coiled coil #status predicted <COI>
F:837-1277/Domain: S2 #status predicted <DS2>
F:1278-1961/Domain: light meromyosin #status predicted <LM>
F:1939-1961/Domain: carboxyl-terminal <CBT>
F:125/Modified site: N6,N6-trimethyllysine (Lys) #status predicted
F:180/Binding site: ATP (Lys) #status predicted
F:694,704/Active site: Cys #status predicted

Query Match 3.2%; Score 279.5; DB 1; Length 1961;
Best Local Similarity 18.8%; Pred. No. 0.00017;
Matches 249; Conservative 223; Mismatches 430; Indels 421; Gaps 60;

Qy 428 RPPKRIHSQEDNVANKQTLASYSRSKATKEDKLLKQEMKS-LAFKAKL-----477
Db 835 KPLLQVSRQBEEMMAKKEELVKREKQLAENRLMELTLOSQMAEKLQLOEQQAETE 894

Qy 478 -----KREKADALEAKK-----EK 492
Db 895 LCAEAEELRLTAKKQLEELIECHDLARVEEERVQHLQAEKKQMQNTQLEEQLEE 954

Qy 493 EDKPKREELKVIIEERLKKKEK-----ERLKVREK-----EREK 530
Db 955 EESARQKLEKVTTEAKLKKLEEQIILEQONCKLAKEKKLEEDRIAETTNLTTESEK 1014

Qy 531----LREKKRY-----VEYLKQWSKPREMECDLKEPPTPVKTRLPETFGDALM 580
Db 1015 SKSLAKLNKHEAMITDLERLRREKQKQLE-----KTRRKLE-----1054

Qy 581 VLEFLNAGELFDLQDEFPD-GVTLEVLAEALVGNDSGLPCCELLFPFLTAIFAQIAEE 639
Db 1055 -----GSTDLSDQIAELQAIKQAKKEE-----LQALARVE 1093

Qy 640 EVAAEQQLTADTKGCSLKS-DLDSCTSLSEILRHILASGADVTSANA---KYRYQKRG 695
Db 1094 EAAQ-----KNMALKKIRELES-QISELQ-----DLESERASRNKAQKQK- 1135

Qy 596 GFDTATDADACMELRLSNLSVKLSSTSVYDITPCKMKILHALCGKLLTVLSTRD-----750
Db 1136 -----DLGELEALKTELEDTLSDTAQQLSRKREQEVNILKLTLEEEAKTHEAQIQE 1189

Qy 751 -----FIDYVDILQAKQEPREL-KAEQH-RKEREAAAARIR-----KRKEELKE 795
Db 1190 MRQHSQAVLEABQLEQTKRVKANLEKAKQTLNENRGELANFVKVLQGDSEHKKK 1249

Qy 796 QEQQKKEKQKLEKDEQ-RNSTAD-LSIGEEREEDFTSIESKDTQEKELDQDMFTED 853
Db 1250 VEAQLQELQVKNFEGERVRLTADKVKTLQVLDNVLTGLSSQSSKSKLTKDFSALBSQ 1309

Qy 854 DPGSHKGRGRKQNGKFKFTROEQINCVTRRELLTADDEEALQKQEQHQRKE-----906
Db 1310 LQDT-----QELLOENRQ-KLSLSLTKLQKQVEDEKNSFREQLSEEEAAKHNLK 1358

Qy	907	-----LLEKIOSAI	ACTNIFPLGRDRMYRYYWIFPSIPGLFIBEDYSGLTEDML	955
		: : : :	: : : :	
Db	1359	QIATLHAQVADMKKMBDSVGCLET	-----ABEVKRC-----LQKDLEGLSQ	1400
		: : : :	: : : :	
Qy	956	LPRSSFNQNVQSDPOVSTKTEGPLEMSESTNIDQPRDHSVOLPKPHKPNRMC	-FYS	1014
		: : : :	: : : :	
Db	1401	-----RHEEKVAAVDKLEKTKT	-----RLQOELDLLVDLD	1442
		: : : :	: : : :	
Qy	1015	SCEOLDOLI	-----EALNSRGHRESALKETLLQEKSRICAQLARFSEE	1058
		: : : :	: : : :	
Db	1443	KQKKFDQLAAEERTISAKYAEERDRAEAREKETALSARALEEAMEQKAELERLNKQ	1502	
		: : : :	: : : :	
Qy	1059	FHFSDKPQPSKPTYSRGRSNAYDPSQMACQLELRDLFLDIEDRIYQOT	-----	1112
		: : : :	: : : :	
Db	1503	FRTEMEDLMSKDDV	--GKSVHELEKSKRALEQQVE--EMKTQLEEELEDEL--QATEDAKLR	1558
		: : : :	: : : :	
Qy	1113	-----LGAIKVT--DRHIWRSALSGRYELLSEENKENGIIKTYNVEDVEEMIDROTK	-----V	1164
		: : : :	: : : :	
Db	1559	LEVNLOAMKAQFERDL	-----QQRDE--QSEKKKQOLVRQVREMEABLEDERKQSRMAV	1610
		: : : :	: : : :	
Qy	1165	IVKORL--LGIKETPTSTVSTNASTPQSVSSVWHYLAAMALFOIEGGIERRRLKAPL	-----	1218
		: : : :	: : : :	
Db	1611	AARKKLEMDLKOLEAHTDSANKNRDEAIKOL	-----RKLOAMKDCMR	1653
		: : : :	: : : :	
Qy	1219	DASDGRSXYKTVLDRWRESLSSASISQVFLHLSTLDRSVIMSKSLNARCKICRKKGDA	1278	
		: : : :	: : : :	
Db	1654	ELDDTRAS	-----REETILAAKENE--KKLSMEAEMIQLQEELAA	1693
		: : : :	: : : :	
Qy	1279	ENMVLDCGDCRGHHTYCVRPKLKTVPGSDWFCPECRPKQRCRLSPRQPSLES	-----D	1333
		: : : :	: : : :	
Db	1694	E-----RAKROAQQRDELADJEANSCKGALALEEKKRLEARIQLE	1736	
		: : : :	: : : :	
Qy	1334	EDVEDSMGGED	-----DEVDGO--BEEGOSBEEVEVEDEDDSOBEEBSVL	1378
		: : : :	: : : :	
Db	1737	EELEEEQGNTELINDRLKKANLQIDQINADNLNLRGHAQKNENARQOLERQKE	-----	1790
		: : : :	: : : :	
Qy	1379	PKGRPOVRLPVKTRGKLSLSSFSRGOQBPGRYPRSQOOSTPKTTVSSKTGSLRKINS	1438	
		: : : :	: : : :	
Db	1791	-----LKVKL	-----QEMEGTVKSKYKASI--TALEAKIAQLEQLDN	1826
		: : : :	: : : :	
Qy	1439	APPTKSLRTASRSTRHSHGLOADVVELLSPPRCRGRCKSANNTPENSPFPNFRVI	1498	
		: : : :	: : : :	
Db	1827	-----ETKERQAACKQVARTKKLKL--DVLLOQVDDERRNNAEQY--DQADKASTLKLQIKRQ	1879	
		: : : :	: : : :	
Qy	1499	ATKSSEQSRSVNTASKLSLOE	-----SESKRRCKRQSPSPSVTLGRR	1542
		: : : :	: : : :	
Db	1980	LEEAEAEARAN--ASRRKLQRELEDATETADAMNREVSSLKNKLRRCDLPFVVPVRMARK	1938	
		: : : :	: : : :	
Qy	1543	SSG	1545	
		: :		
Db	1939	GAG	1941	
		: :		

RESULT 29

RESULT
T16270

116270
hypothetical protein F35D11.11 - *Caenorhabditis elegans*

C:Species: *Caenorhabditis elegans*

C:\Species: CACONINABUTIS CREGANS
C:\Date: 20-Sep-1999 #sequence revision 20-Sep-1999 #text change 09-Jul-2004

C;Accession: T16270

R;Fulton, B.

submitted to the EMBL Data Library, June 1995

A;Description: The sequence of *C. elegans* cosmid F35D11.

A;Reference number: Z18487

A;Accession: T16270

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-1827 <FUL>

A;Cross-references: UNIPROT:Q20042; EMBL:U29381; NID:g868214; PID:g868224; PIDN:AAA68757

A;Experimental source: strain Bristol N2

C;Genetics:

A;Gene: CESP:F35D11.11

A;Intron8: 76/2; 131/3; 159/3; 185/3; 221/3; 253/3; 320/1; 869/3; 1133/3; 1205/2; 1250/1


```
Qy 1090 EKQELRLRDLFDLIEDRIYQGTLCGAIKVTDRIHWSALESGRYELLSEENKENGCIITV 1149
Db 914 -RDLDDKLRLKMWELTDEE-----KEEDRKKGQKTLNEERMKLM-EQKEEAMLVATK 961
Qy 1150 NEDVEEMIDEQTKVI-----VKDR-----LLGIKTEPTSTVSTNASTPQSVSS 1193
Db 962 HATT-----IDQOTRISVLGDEVEKLTAGIAERESSINALSENWELISKLETTAEALEK 1017
Qy 1194 VVHYLAMALFOI-----EQIERRF---LKAPLDASDGRSYKTVTLDRWRRESLLSSASL 1244
Db 1018 LKDELAVMLKQNSLKNKGKGLSEKWNBERKKIQDLADQLREANKVHNMR---MKVNL 1074
Qy 1245 SQVFLHLSTLDRSVI-WKSILNARCKICRKKGDAENNVLCGDCDRGHHTYCVRPKLTIV 1303
Db 1075 EE---KKNELDNQVTLNKRQRLBIQLMDKA--AKNEVSGDLURKMEHD--AQSMLKQA 1127
Qy 1304 PEGDWFCPCRPKORC-----RRL-----SPRQPSLEDEDEVEDSMGGEDDEV 1347
Db 1128 QNEQFRLTDLKVKRALQDQENQRLVNDLATVKAPEVRETSKSAISDILDKYSAEKA 1187
Qy 1348 DGDEBEGQS-----PEEYVEQDED---DSQEE-EVY-----SLPKRGRPOV 1386
Db 1188 NKGELDNQRLSDLATVTLKLERQELKAKDSNRLRDSQKRFEEVQSKLANLQKSAVESL 1247
Qy 1387 RLPVKTRKGLSSP-----SSRQOQEPGRYPSRS-----QQSTPKTIV 1425
Db 1248 QNPMSNSGRNRSIYVDIPRAASSITGLNENSDVPLRSSPSVRFADSQNNQRAVDSMDV 1307
Qy 1426 SSKTGRSLR-----KINSAPPTKSLRTASRSTRSHGFLQADVFVLLSPRRKRG 1478
Db 1308 SSSGVTVLRFKRIEQLQEANADLSDALEKAKDELORNEKL-ADROMVIERVERQLVH 1366
Qy 1479 RKSANVTPEPNFPNFRVIATKSEQSRVNIA-----KLSQESSEK 1523
Db 1367 ITEERTNIENMTSQROMYLTNEESSRSREHIRMKARISTLEHLAEKSK 1419

RESULT 30
A:59287
myosin heavy chain - fluke (Schistosoma mansoni) (strain Brazilian LE)
C:Species: Schistosoma mansoni
C:Date: 09-Jun-2000 #sequence_revision 09-Jun-2000 #text_change 09-Jul-2004
C:Accession: A59287
R:Weston, D.S.; Schmitz, J.; Kemp, M.; Kunz, W.
Mol. Biochem. Parasitol. 58, 161-164, 1993
A:Title: Cloning and sequence characterization of a complete myosin heavy chain cDNA from
A:Reference number: A59287; MUID:93211444; PMID:8459827
A:Accession: A59287
A>Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-1940 <MES>
A:Cross-references: UNIPROT:Q02456; GB:L01634; PIDN:AAA29905.1
A:Experimental source: strain Brazilian LE
C:Genetics:
A:Gene: MYH
C:Superfamily: myosin heavy chain; myosin motor domain homology
F:82-752/Domain: myosin motor domain homology <NMO>

Query Match 3.2%; Score 279; DB 2; Length 1940;
Best Local Similarity 18.5%; Fred. No. 0.00017;
Matches 294; Conservative 232; Mismatches 525; Indels 542; Gaps 64;

Qy 206 VTGRPG-----LTYQEALESEKAKQNLQS-----FPEPLIIPVLVL 242
Db 618 VVGEPKTKGSGFLTVTYMHR-ESLNLKMLKNLQSTSPSFIRCIVNPFKQPGVIDAHLV 676
Qy 243 TSLTHRSRLHE--ICDD-----IFAYVDRYFVEETVVRNNGARLQCTILEVLPSPS 293
Db 677 LHQLHNGVLEGIRICRKGFPNRMYSBFQRY-----SILAPN-----VIP-- 718
Qy 294 HQNGFANGHVNVSVDGETIIISDSDSETQSCSFQNGKKKDAIDPLLFKVKVQPKLHE 353
Db 719 --DGFVDGR-----QVTEK----- 730
```

```
Qy 354 SAIKVAQISRRKHLFRSDKULFLK-----QHCE-----POEGVIKIKASSLTVYKIAEQ 404
Db 731 --ILEATQLD--KNLYQCGNTKVFVKAGTLAHLDRDDKNGIISLFOAIRGYLMRQ 786
Qy 405 DFSYFFPDPTTFIFSPANR-----RGRPPKRIHSQED--NVANKQTLASYRSKATKE 457
Db 787 YKK--LQQRVALTLMQRIKYLVLRNWPNWRLYTKVKPMLNIAR----- 830
Qy 458 RDKLLKQEMKSLAFKAKLRERKADALEAKKCKEKEDKEKREELKKTIVEBERLKKEBK 517
Db 831 -----QEEMKKAABELAKEE-----YEKLEKLKLEEQNVTVLQOK 870
Qy 518 ERLKVEREKREKREKRYVEYLKQSKPREMECDLKLPEP-----TPV 566
Db 871 NDLFLQLQTEQDSLADAEKISKVLQ---RGDWE-ORIKELERLADEEDQANLTV 925
Qy 567 KTRLPPIFGALMVLEFLNAGFLFDLODFPDGVTLEVLVEALVGNDSBGPLCELLPF 626
Db 926 KKKVSAEI-----BELKK--DVEDLES-----SLQKAEQEKQTKDNQ----- 960
Qy 627 FLTAIFOAIAEBEEBEVAK-----EQLTDAATKGCSLKSL--DLDSCTLS 668
Db 961 -IRTLQSEMAQODEMI GKLNKDKNLEBEQNKRTQEQALQAEEDKVNHLNKLKAKLES--TLD 1018
Qy 669 EILRLHLASGADVTSANAKYRYQKGGFDATDDACMELRLSNFSLVKLSTSVYDLTP 728
Db 1019 EM--EENLARBEQKIRGDEVKSKRLEGDLKATQETVDDLERVKGRLBQLR----- 1067
Qy 729 GEKMKILHALCGKLLTLVSTDFIEDYVILROAKQEFRELKAEQHRKEREANAIRKR 788
Db 1068 -RKEAETGGLSGK-----FEDEGLVAQLQRIKELQTRIQELEEDLEAERAARS 1116
Qy 789 KEELKKEQ-EOKMKEKQELKEDQORNSTADISIGEEREDFTSIESKTEQEKLDQDM 847
Db 1117 KAEKSRQOLESELEEVVDLEE-----QOGATAAQSDLTKKRAELMK 1159
Qy 848 FTEDEDDPGSHKRRGRKQNGPKFTRQOINCVTRRELTATDEEEALKQEHQREKEL 907
Db 1160 LKRDLED-----TRLQN-----EQAIATMRKKQSDAINELADQLDQANKAKAK 1202
Qy 908 LEKIQSAIACTNIFPLGRDRMYRYWIFPSPGLFIEDYSGLTEDMLLPPRSSFQNVV- 966
Db 1203 AEKERSQFKA-----ELDDAHNQVDSIMKAKLNSEKTVK 1236
Qy 967 --QSQDPQVSTKTEPLMSESTNIDQ--GPRDHSVOLPKPVHPNRCFYSSCSQLOQL 1022
Db 1237 ALESQLEVSVK-----LDEATRLNLEQASTKARSSQEVSELQR-----QLEEAESQ 1283
Qy 1023 IEALNSRGHRESALKETL---LOEKSRICAOL-----ARFSEKPFHPSD- 1063
Db 1284 LSQNLKIKQQLSAQLEEARHSLEDES RMKAKLNGEVRLNLTSDLSLRETLEEQSAKGDL 1343
Qy 1064 -----KPQDPSKPTYSRG-----RSSNAYDPS-----QMCSE----- 1090
Db 1344 QRQLKQLGELQQLRSRGGGGVRSBEVEBELKRMNAKIPALESEASAKSKCQLEKT 1403
Qy 1091 -KQLELRRLRDFLLDIE-----DRIYQGTLCGAIKVTDRIHWS 1126
Db 1404 KARLQGELEDUMVDVERANGLASQLERKQNNFNRTLAEWQKYQADSQLENAQDARGQ 1463
Qy 1127 ALESGRYELLSE-----NKENGIIKTVNEDVEE-----EIDEQTKVIV 1166
Db 1464 STEIFRLKAQLEEVHEQMEGLRREN--KNLSDEIHDITQELGEGRSVHEIDKNRR--- 1517
Qy 1167 KDRLLGKTETPTSTVSTNASTPQSVSVVHYLALFOIEQIGIERRFLKAPLDASDGRS 1226
Db 1518 --RLEMEKEELQAALEAEASALEQEAQVQRAQLEMSQIROEIDRLRAKEEKEEFATRKN 1575
Qy 1227 YKTVLDRWRRESLLSSASLSQVFLHSLTLDRSVINSKILNARCKICRKKGDAENM----- 1281
Db 1576 HQAMESQQASLEAEG-----KGKAEAMRVKKK 1603
```

Qy 1282 -----VLDCGDCRG-----HHTYCVRPKL 1300
Db 1604 LEQDINELEVLSDGNARARAEQENVKFOQVRELQSOLEDQORQDRLLEQQAERR 1663
Qy 1301 KTVPEGWFCPECR-----PKQCRRLSRFRQPSLESDEVE-----DSMGEDDEVGDGE 1351
Db 1664 ATVLAGE--LDLEALDQASRSKIAEABAE--ASDRATEMSQTASLAQAQKLEADL 1720
Qy 1352 EEGQSEEEVEVEQEDDS-----QEEVEVSLPKGRPOVRLPVK--- 1391
Db 1721 AAMQADLEEAANEAKQADERAKKAWADARSVEFIRQEQHTQHVAKRQKLEIQVKELM 1780
Qy 1392 TRGLSSFSRSGOQEGRPYSQOSTPTKTVSSK-----TGRSLRKINSAPPTETKSL 1447
Db 1781 ARLESESGAMNGKANGKLEQVRLELETLAEQRHGTQKNLKYD-----RRMKEI 1836
Qy 1448 RIASRSTRHSHGPIQADVVELLSPPRRKRGKSNANTPENSPPNFRVRIATKSSSQSR 1507
Db 1837 SLQAEEDKSHDMQE--LVEKLQGIKITYKRVQVQEAEEIAAINLAKYKIOHIEDAE 1894
Qy 1508 SVNTASKLSLOESKRCRKRQSPSPVTLG 1540
Db 1895 RADQAEQ-ALQKRAKRNSSVSTARGVSPAPLG 1926

RESULT 31
QRMSP1
microtubule-associated protein MAP1B - mouse
N;Alternate names: microtubule-associated protein MAP1(X); microtubule-associated protein
C;Species: Mus musculus (house mouse)
C;Date: 30-Sep-1991 #sequence revision 30-Sep-1991 #text_change 09-Jul-2004
C;Accession: S07549; S44387; A33645
R;Noble, M.; Lewis, S.A.; Cowan, N.J.
J. Cell Biol. 109, 3367-3376, 1989
A;Title: The microtubule binding domain of microtubule-associated protein MAP1B contains
A;Reference number: A33645; MUID:90094539; PMID:2480963
A;Accession: S07549
A;Molecule type: MENA
A;Residues: 1-2464 <NOB>
A;Cross-references: UNIPROT:P14873; EMBL:X51396; NID:952999; PIDN:CAA35761.1; PID:953000
R;Sanchez, C.; Padilla, R.; Paciucci, R.; Zabala, J.C.; Avila, J.
Arch. Biochem. Biophys. 310, 428-432, 1994
A;Title: Binding of heat-shock protein 70 (hsp70) to tubulin.
A;Reference number: S44387; MUID:94234720; PMID:8179328
A;Accession: S44387
A;Status: preliminary
A;Molecule type: protein
A;Residues: 653-663,'IC' <SAN>
C;Superfamily: microtubule-associated protein MAP1B
C;Keywords: microtubule binding; phosphoprotein; tandem repeat
F;589-786/domain: microtubule binding #status experimental <MTB>
F;589-592, 639-642, 649-652, 655-658, 660-663, 668-671, 674-677, 679-682, 683-686, 687-690, 691-69
R-K-E/D-X)

Query Match 3.2%; Score 278.5; DB 1; Length 2464;
Best Local Similarity 18.8%; Pred. No. 0.00024;
Matches 299; Conservative 243; Mismatches 484; Indels 563; Gaps 74;
Qy 171 DEEVFYCKVNE-----IFRHVDDPFERILCNLSVWCAVTCRPGLTQEALESSEKK 223
Db 379 BEACFTLQYLNKLSMKPPLFRSGNTIEPILFQRM-----GVGKLEMYVLNVPVKSKE 433
Qy 224 ARQNLQSPF-----BPLIIPVLYLSLTHSRSLH-----ETCDDIFAYVK 263
Db 434 MQYFMQQTGNTKDAEAILLNGEVDIPSYLTSVSSLIWHHPANPAEKILRVLPQNS 493
Qy 264 DRYFVEETVEIRNNGARLQCTILEVL--PFSHQNGFANGHNSVDGETIIISDDSET 321
Db 494 TOYNILEGLEKLKH-----LDLKLQPLATQKDL-TGQVPTPPVKVQLKQRADSR 543

Qy 322 QSCSFQNGKKDAIDPLLFYKVQPTKKELHESAIVKATOISRRKHLFSRDKLKLFLKQH 381
Db 544 -----SLKPATPKVASKVRKESKEETPB--VTTSQVEKTPKVESKEV---LVKK 590
Qy 382 CEPOGVIKIKASSLSTYKIAEQDFSYFFPDPPTTFIFSPANRRRGRPPKRIHISQEDNV 441
Db 591 DKP-----VKTESKPS--VTEKVS-----SKEEQSPVKA-----EV 620
Qy 442 ANKOTLASYSKATKERRDKLLKQEMKSLAFKAKLREKADALEAKKE-----KEDKE 496
Db 621 AEKQATES-RPKVTK--DKVVK-KEIKT-----KLEEKKEPKKEVVVKEDKTKPLKQDEK 672
Qy 497 KREELAKIVVEERLKKKEERLKVREKE-----REKLREKRYKVEYLKOWSKPRE 550
Db 673 PRKEEVKGIKKE--IKKEERKELKEVKETPLKDAKEVKKEKKEVK--KEKEPKK 728
Qy 551 DMECDLKLPEPTPVKTRLPPEIFGDALMVLEFNAFGLFDLQDEFPDGGVTTLEVEEA 610
Db 729 EIK-KISKDIKKSTP-----
Qy 611 LVGNDSGPICELLFFPLTAIFQAIABEEBEVAKEQLTDATKCSLSKSLDLDSTLSEI 670
Db 743 --QSDTKK-----SALKPKVAKKESSTKKEPLAAGKLKD----- 775
Qy 671 LRLHILASGADVTSANAKYRQKRGFPDADDACMELRLSNPSLVKKLSSTSVYDLTPGE 730
Db 776 -----KGVKVIKKEG--KTEEA-----ATAVGTAAITAAVVAAG-- 810
Qy 731 KMKILHALCKGLTLVSTRDFIEDYVDILRQAKQEFRELKAEQHRKEREEREAARIRKKE 790
Db 811 -----IAASGPVKELEAERSLMSPPDLTK---DFEELKAE-----IDVAKDIKPQLE 856
Qy 791 EKLKEQKQKKEKQEK-----LKDEORNSTAD-----ISIGEREREDFTSIESKTEQ 840
Db 857 --LTIESEKLKETQPGYAYVIQETVSKGSAESPDEGITTTTEGECEQTPPELEPVEK 914
Qy 841 KELDDMFTDEDDPGSHKGRGKRGQNGFKETROEQINCVTRELLTADDEERALKQEH 900
Db 915 QGVDD--IKFDEGA-----GFEESE-----TGDEYKAETE-- 946
Qy 901 QRKEKELLEKIQSAIACTNIFPLGRDRMYRYWIFPISPLFIEDYSGLTEDMLPRPS 960
Db 947 -----
Qy 961 SFQNVOSQDPQVSTKTGEPMLSESTSNIDGPRDHSVQLPKPVH-KPNRCWCFYSSCEQL 1019
Db 947 -----EAEPE-----EDGEDNASGSASKHSPTEDDESAKAEADVHLKRESVVSGDDRA 997
Qy 1020 DOLLEALNSGRHSALKETLLQESKRICAQLARFSEKPHFSDKPOPDSPKPTYSRGRSS 1079
Db 998 EEDMDDVLEKGEAQ-----SEEEGEEDKAE-DAR-----E 1028
Qy 1080 NAYDPSQWCAEKQLELRDLFLDIEDRIYQGTLGAIKVTDRHIWRSALESG-----RYEL 1135
Db 1029 EGYEPDKTEAE-----DYMAVADK-----AAEAGVTEEOVGY 1061
Qy 1136 LSEENKENGIIKTVNEDEVMEIDEQTKVIVKDRLLGIKTKETPTVS---TNASTPQSVS 1192
Db 1062 LGTSAKQGIQSPSRSEPASS-----IHDSTLPGSESEATASDENREDOPEEFT 1111
Qy 1193 SVVHYLAWALFQIEQHTERRFLKAPLDASDSGRSYKTVLDWRRESLLSSASLSQVFLHLS 1252
Db 1112 ATSGY-----TGSTIEISSEPTMDMSTPRD--VMSDETNN--ETESPSSEFVNIT 1160
Qy 1253 TLDRSVWKSILNARCKICRKKGAENMVLCDGCDRGHHTYCVRPKLKTVPEGDWFOPE 1312
Db 1161 KYESS-LYSQEY-----
Qy 1313 CRPKQRCRLSRFRQPSLESDEVEDSMGDEDDVDGDEBEGQS-----EEEEVEV 1363
Db 1172 -----SKPAVAFNGL--SEGSKTDATDGKYNASASTISPPSSMEEDKPSK 1216

QY 1364 EQEDDSQBEVEVSLPKRGPQVRLPVKTRGKLSLSSFSRGGQOBERYPSPSQSTPKT 1423
 Db 1217 SALRDAYCSEE-----KELKASAELEK-----DVSDRLSPAKSPSL--PSPSPS 1260
 QY 1424 TVSSKTGRSLRKIN-SAPPTETK-SURIASRSTRSHGFLQADV-----FVELLS 1471
 Db 1261 PI-EXTPLGERSVNSFLTNEIKVSAEGARSV--SPGVQAVVEHCASPREKTLVVS 1317
 QY 1472 PRKRGGRKSNANP-ENSPNPNFRVIATKSEOSRSVNIASKLSQSESKRR-----1525
 Db 1318 PQSVVTG--SAGHTPYQSPTEKSHLPTVSENAQAVPVSFSEAKDENERSASLSPM 1375
 QY 1526 -----CRKQSPSPVTLGRSSRGQGVHE--LSAFEQVLVVLVRHDSWPPL 1573
 Db 1376 DEVPDSESFVEKVLSPURSPPLGSESP-----YEDFLSADSKV--LGRRSSES-PPE 1425
 QY 1574 KLVSKIQVDDYDIKPIALMIIEKVNKCYKLASEFIDIELMFSCFYNP-RNYS 1632
 Db 1426 GNGKQGFDPDRSPVSDLTSTGLYQDK-----QEEKSTGFIPKE-----DFGPEKTS 1474
 QY 1633 EAKAGTRIQAFPHIQAKLGLHVTSPNVD 1661
 Db 1475 DVETMSS-QSALALDERKLGDSVPTQID 1502

RESULT 32
 S03166
 myosin heavy chain, gizzard smooth muscle (similarity) - chicken
 N:Contains: myosin ATPase (EC 3.6.4.1)
 C:Species: Gallus gallus (chicken)
 C>Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 09-Jul-2004
 C:Accession: S03166; A27066; A26045; A36604; A43298
 R:Yanagisawa, M.; Hamada, Y.; Katsuragawa, Y.; Imamura, M.; Mikawa, T.; Masaki, T.
 J. Mol. Biol. 198, 143-157, 1987
 A>Title: Complete primary structure of vertebrate smooth muscle myosin heavy chain deduced
 A:Reference number: S03166; MUID:88118918; PMID:2892941
 A:Accession: S03166
 A:Molecule type: mRNA
 A:Residues: 1-1979 <YAN>
 A:Cross-references: UNIPROT:P10587; EMBL:X06546; NID:G63633; PIDN:CAA29793.1; PID:G63634
 A>Note: part of this sequence was confirmed by protein sequencing
 R:Maita, T.; Onishi, H.; Yajima, E.; Matsuda, G.
 J. Biochem. 102, 133-145, 1987
 A>Title: Amino acid sequence of the amino-terminal 24 kDa fragment of the heavy chain of
 A:Reference number: A27066; MUID:88032919; PMID:3312184
 A:Accession: A27066
 A:Molecule type: protein
 A:Residues: 2, 2', 4-204 <MAI>
 R:Onishi, H.; Maita, T.; Miyaniishi, T.; Watanabe, S.; Matsuda, G.
 J. Biochem. 100, 1433-1447, 1986
 A>Title: Amino acid sequence of the 203-residue fragment
 A:Reference number: A26045; MUID:87194651; PMID:3571180
 A:Accession: A26045
 A:Molecule type: protein
 A:Residues: 653-855 <ONI>
 R:Onishi, H.; Maita, T.; Matsuda, G.; Fujiwara, K.
 J. Biol. Chem. 265, 19362-19368, 1990
 A>Title: Lys-65 and Glu-168 are the residues for carbodiimide-catalyzed cross-linking be
 A:Reference number: A36604; MUID:91035476; PMID:1977747
 A:Accession: A36604
 A>Status: preliminary
 A:Molecule type: protein
 A:Residues: 54-67;146-183 <ON2>
 R:Cole, D.G.; Yount, R.G.
 Biochemistry 31, 6186-6192, 1992
 A>Title: Stability and photochemical properties of vanadate-trapped nucleotide complexes
 A:Reference number: A43298; MUID:92329440; PMID:1385724
 A:Accession: A43298
 A>Status: preliminary
 A:Molecule type: protein
 A:Residues: 169-183 <COL>
 A:Superfamily: myosin heavy chain; myosin motor domain homology
 C:Keywords: actin binding; ATP; blocked amino end; coiled coil; hydrolase; methylated am

F:87-777/Domain: myosin motor domain homology <MMOT>
 F:177-184/Region: nucleotide-binding motif A (p-loop)
 F:565-578/Region: actin binding #status predicted
 F:639-653/Region: actin binding #status predicted
 F:850-1940/Domain: actin binding #status predicted
 F:850-1290/Region: S2
 F:1291-1979/Region: light meromyosin
 F:1941-1979/Domain: carboxyl-terminal <CBT>
 F:2/Modified site: blocked amino end (Ser) (in mature form) #status experimental
 F:128/Modified site: N6,N6-trimethyllysine (Lys) #status experimental
 F:183/Binding site: ATP (Lys) #status predicted
 F:707,717/Active site: Cys #status predicted

Query Match 3.2%; Score 278; DB 1; Length 1979;
 Best Local Similarity 18.4%; Pred. No. 0.00019;
 Matches 244; Conservative 226; Mismatches 445; Indels 414; Gaps 54;

QY 212 LTYQALSEKAKARQLQSPPEPLIPVLYLTSLTHRSPLHEICDIF-----AYVKD 264
 Db 959 LDLEQLSEEAARQLQ-----LEKVTADGKIKKXMEDDILIMEDQNNKLTKE 1006
 QY 265 RYFVEETVEIRNNGARLQCTILEVLPSSHQNGFANGHVNSVDGETIIISDSDSDSTQSC 324
 Db 1007 RKLLEERSVSLTNNLAE-----EERKAKNUTKLNKH-----ESMI-----SELEVR 1048
 QY 325 SPQNGKKXDAIDPLLFYKVPQTKKELHESAIKATQISRRKHLFSRDKL-----FLK 379
 Db 1049 LKKEEKSRQLEKI--KKLEGESDLHEQIAELQAQIAELAKKAEELQAALARLE 1106
 QY 380 QHCPEQGVKIKASSLSYTKIAEQDFSYFFPDPTTFISPANRRGRPPRIHLS--Q 437
 Db 1107 DTSQKNALK-----KIRELE-----SHISDLQ 1130
 QY 438 EDNVANKQTLASYRSKATKERRDKLKQEMKSLAFKAKLREKADALEAKKKEKDKK 497
 Db 1131 EDLESEK-----AARNKAEKQK-----RDLSELEALKTELEDTLTATTQQLRAK 1177
 QY 498 KRBE--LKKIVEEE-----RLKKK-----EERLKYVEREKEREKRE 533
 Db 1178 REQVTVLKRALREETRTHAEQVEMRQKHTQAVEELTEQLQFQKRAKANLDTKQTLEK 1237
 QY 534 EKRKYVEVLKQWSPREDMECDLDELPEPTPVKTRLPPEIFGDALMWLEFLNAGELFD 593
 Db 1238 DNADIANEIRLSQAKQDVEHKKK-----LEV-----QLQD 1269
 QY 594 LQDEPPDG-----VTLVLEALVGNDSGFLPCCELLFFFLTAIFAQIAEERE 640
 Db 1270 LQSKYSDGERVTELTNEKVKHLQIEVNTSLNEAESKNILKTDVATLGSQ--LQDTQ 1327
 QY 641 EYAKEQLTDADTKGCSLSDLDSCSLSEILRLHLTA-----SGADVTSANAKRYQ 692
 Db 1328 ELQGETRQLNVTTKLQLEDDKNSLQELDEVEAKQNLERHISTLTILQSDSKKLQ 1387
 QY 693 KRGGFDAT-----DDACMEIRLSNPPLVKLSTSV-YDLTPGEKMKILHALCGKLLTVS 747
 Db 1388 E---FTATVETMEEGKKLQREIESLTQQFEKAAASYDKLEKTKRLRQLOELDLVDLDN 1444
 QY 748 TRDFTEDVDVILRQAKQFRELKAEQ-----HRKEREAAAIRKREKLEKEQEKWK 801
 Db 1445 QRLVSN-----LEKQKKPQDMLAEKKNISSKYADERDRAEAERE-KETKALSARALE 1499
 QY 802 EKQEKLEDEQORNSTADISIGEEEREDFTSTESKDTBQKELQDMFTEDDEDPGSHKRG 861
 Db 1500 EALAEKKELERTNKL-----KAEMEDL--VSSKDDVGKNVH-----L 1536
 QY 862 RRGKRGQNGKFEFTREQINCVTRELTFADBEELAKQEHQKKEKLEKIKQSAIACTNIF 921
 Db 1537 EKSKR-----TLEQQV-----EEMKTOLEBELELQAA-----1564
 QY 922 PLGRDRMYRYWIPPSIFGLFIEDYSGLTEDMLLPRSSFNQNVQSQDPQVSTKTGEP 981
 Db 1565 -----EDAKURLEVNMQAMKSQFERDLOARDEQNEEKRR-----1598

Qy	982	MSESTSNIDGPRDHSVQLPKPVHKPNRWCYSSCEQLDOLIEALNSRGHRESALKETLL	1041
Db	1599	-----QLLKQLHE-----HETEDEL-----	1613
Qy	1042	QEKSRICAQLARFSEEFHSDPOPDSPKTYTSGRGSNAYDSQWCAEQLELRDLDFL	1101
Db	1614	ERKQALAAAK-----KLEVDVKOLESQVDSANKAREEAIKQRLKLOAQWKDYQ	1664
Qy	1102	LDIEDRIYOGTLGAIKVTDTRHWRSALESGRYELL--SEENKENGIIKTWNEDVEEMEID	1159
Db	1665	RLDD-----ARAAREIFATARENEK-----KAKULEAEIIQLQ	1699
Qy	1160	EQTQKVIKDRLLG--IKTETPTSTVSTNASPPQSVSVVHYLAMALFQIEGIERRFLKAP	1217
Db	1700	EDLAAAEARKQADLEKEEMAEELASANSRTSLQDEKRRLEARIAQLEELD-----	1752
Qy	1218	LDASDSGRSYKTVLDRWRESLLSASLSQVFLHLSITLDRSVIWSKSIILNARCKI	1277
Db	1753	-----EHSNIETWSDRMKAVQAQELNN---ELAT-ERATAQKNE--NARQOLESQNK	1802
Qy	1278	AENMVL--CDGCDRGHHTYCVRPKLKTVPEG-----DWFCPECRPKQCRRLSFROR	1327
Db	1803	LRSLQWEG-----AVKSFKSTIALEAKIASLEOLEQAREKQAAAK--TLRQK	1853
Qy	1328	PSLESDEVEDS--MGGEDDVGDEEGQSEEEYEVEQDEDSQEEVEVSLPKRGRPQV	1386
Db	1854	-----DKKLKALLQVEDERKQAEQYKQAEKGNLRLKQLKRLQLEAEES-----Q	1900
Qy	1387	RLPVKTRGKLSSSFSSRGQOEP--GRYPSRSQOSTPKTTVSSKTGSLRKSINAPTEYK	1445
Db	1901	RINANRR--KLQRELDEATESNDALGR-----EVAALKSLRRGNPEVPSPRRSG	1950
Qy	1446	SLRIASRST	1454
Db	1951	GRVVENAT	1959
RESULT 33			
T20532			
hypothetical protein F07A11.6b - Caenorhabditis elegans			
C/Species: Caenorhabditis elegans			
C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004			
C/Accession: T20532; T27777			
R;Palmer, S.			
submitted to the EMBL Data Library, October 1995			
A/Reference number: Z19287			
A/Accession: T20532			
A/Status: preliminary; translated from GB/EMBL/DBJ			
A/Molecule type: DNA			
A/Residues: 1-2722 <W1>			
A/Cross-references: UNIPROT.Q15135; EMBL.Z66511; PIDN.CAB54211.1; GSPDB:GN00020; CESP:F07A11.6b			
A/Experimental source: clone F07A11			
R;Gajadaty, S.			
submitted to the EMBL Data Library, March 1996			
A/Reference number: Z20417			
A/Accession: T27777			
A/Status: preliminary; translated from GB/EMBL/DBJ			
A/Molecule type: DNA			
A/Residues: 1-2722 <W12>			
A/Cross-references: EMBL.Z69904; PIDN.CAB54502.1; GSPDB:GN00020; CESP:F07A11.6b			
A/Experimental source: clone ZK20			
C/Genetics:			
A/Gene: CESP:F07A11.6b			
A/Map position: 2			
A/Introns: 36/2; 92/2; 182/2; 272/3; 344/2; 426/2; 490/1; 541/2; 700/3; 770/3; 1286/3; 1			
Query Match			
3.2%; Score 277.5; DB 2; Length 2722;			
Best Local Similarity 19.4%; Pred. No. 0.0003;			
Matches 310; Conservative 203; Mismatches 536; Indels 547; Gaps 66;			
Qy	203	SCAVTGRPGTYQEALESEKARQNLQSFPEPIIIPVLYLTSLTHRSRLHEICDDIFA--	260
Db	775	SMSTGRPA-----SIOTLRHOSVWFPEDVSIPPP-TH-----DENMAPR	817

Qy	261	YVKDRYVFVEFVIRNN--GABLQCTI-LEVLPSPHQNGFANGHVNSVDGTEIILSDSD	317
Db	818	GTPPSRSSSTWPLASPPPGTQIONLLTPIVPPHLLIAATSTGTHSVSSSAHSTPRHS	877
Qy	318	DSETO-SCSFONGKKKXDAIDPLLFVKYVQPTKKELHESAI-----VKATOISRRKHL	368
Db	878	ISGTPVHCPSNSKTSQPPTP-----KSRPEKVOIRHDTISKSGPSNAINALO-ARQSQM	931
Qy	369	FSDKCLKLFLKQCHCEPOEGVIKIKASSLSYKIAEQDFSVFFDDPPTFIFSPANRRGR	428
Db	932	TSGDPK-----SASTPVRDAGDULVAQINSQ-----PNLGRKL	969
Qy	429	PPKRIIHSQBDNVANKQTILASYRSK-----ATKERDKLLQOEMKSLAPEKAKLKREKAD	483
Db	970	PRIEKKSSALQIONHQPHPSNANSTPSTPSTHQAMFKDKERKKKKEKEERERE	1029
Qy	484	ALBAKKKEKEDKKKREBK--KIVBEERLKKKEERKLKVEREKEREKL-----BEKK	537
Db	1030	RREMKRETBEENKGMERAKRLEDERQERKKERKEDERDERKEKVKRKAEEKL	1089
Qy	538	YVEYLKOWSKPRDMCEDDLKELPEPTPVKTRLPPFIFGDALMVLEFLNAFGLFDLQDE	597
Db	1090	KKHGRGSDSDSDNSDELDOVRKSYK-----	1120
Qy	598	FPDGVTLEVLFEALVGNDSGLCELLFFLTAFQIAAESEEEVAKEQ-----LTDADT	652
Db	1121	-----MTOEEKDHQALLLSKGIIENLKSRRSDKEAHDSEFKWQKQSQORRVLISSDD	1176
Qy	653	KGCSLKLSDLDSCTLSEILRLHLASGADVTSANAKYVQKRGFPATDDACMELRLSNP	712
Db	1177	EG-----GKGDKGNS-----SNGEESDSKADLPPPPAPP	1207
Qy	713	SLVKLSSTSVYDLTPGEKMKILHALCGKLLTLVSTDFIEDVVDILRAQAKOFRELKAE	772
Db	1208	SLSE-----SADQRLVKLREKSGELTTSDDENDHNDAGEIHOQRLTREDNRKR	1257
Qy	773	Q-----HRKERBEAARIKRKEE-----KLKEQEQKMKQKELKXEDQORNSTADI	819
Db	1258	QKSLTAYSDQGERKNVPKWRDDSDAAAGHGWSAKDQKQKRLKEHRS-----	1313
Qy	820	SIGEEERDDPTISKQTEQKL-----QDMTEDEDDPGSHKRGGRKRGQNGKPEFT	875
Db	1314	-----ED-----ESKNAKRDFRDIPIHEDVSEDEBEDGSRSR-----	1346
Qy	876	RQEQNCVTRRELTABEEALKQEHQKKELEKIQSALACTINIFPLGRDMYRYWIF	935
Db	1347	RQSTSTISN-VTAK-----RKEX-----	1365
Qy	936	PSIPGLFIEDYSGLTEDMLPRPSSFQNNVQSDPQVSTKTCGEPLMSBSTSNIDQPRD	995
Db	1366	-----SGKTPRLRVPEP-----YCTPLLSP-----	1385
Qy	996	HSVQLPKPVHPNRWCIFYSCQOLDOLIEALNSRGHRESALK--ETLLOEKS-----ICA	1049
Db	1386	---KILSPKH-----LSPKTSSTSKRSSIDHENLISPRQNRRTTSST	1426
Qy	1050	QLARFEEKHFSDKQPDSPKTPYSGRSSNADPS-----QMAEKQLELRDLFULL	1102
Db	1427	STATTSKKEALSIPKPLSPPPVYATKSSVSSIDDPISIRDEFMSNAADSPMSTTGPMVL	1486
Qy	1103	DIEDRIYQGTGLKAVTDRIHWESALESRGYELLSENKENGII--KTVNEDVEEMEIDE	1160
Db	1487	T-----KAAMKAFNSTPPKKNS--SGQHDSSSSSSSDSDSDSDSDSDSDDEVPK	1538
Qy	1161	QTKVIVKDRLLGIKETPSTVSTNASTPOSVSVHYLAWALFQIEQGIERRFLKAPLDA	1220
Db	1539	QTEPV-----TSIPVVASDNGS-PENV-----VVEPSPIV	1567
Qy	1221	SDSGRSYK--TVLDRWRESLLSASISQVFLHLSTLDRSVNSKSLNAECKICRKKGDA	1278
Db	1568	SQTPREPPETIIEQSES-----	1586

Db 1721 KEELRLLEAR---IAQLBEELEEEQNTLINDRLKANKLQIDQINADNLNLERGHAQVE 1777
Qy 1361 YEVEQEDDDSQEEEVSLPKGRPOVRLPVKTRGKLSSFSRGOQPGPYPSRQOST 1420
Db 1778 NAROQLERQNE---LKVKL-----QEMEGTVKSKYKASI 1809
Qy 1421 PKTVSSKTGRSLRKINSAPPTETKSLRIASRSTRSHGHPLOADVVELLSPRKRGRK 1480
Db 1810 ---TALEAKIAQLEBQDN---ETKERAQACKVARTKCLK-DVLLQVDDERRNAEYK 1862
Qy 1481 SANNTPENSPNPNFRVIATKSSSEQRSVNIASKLSLQ-----SESKR 1524
Db 1863 --DQADKASTRLKQLKRQLEAEBAEBAQAN-ASRRKQLEADATATADAMNREVSSLN 1919
Qy 1525 RCRKQSPPEPVILGRSSRGQGVHLSAFELVVELVRHDDSWPFLKLVSKIQVDPY 1584
Db 1920 KLRGDLFPVTVRLVRK-----GTLELS-----DDDDSKASLINETQPPQC 1962
Qy 1595 YD-----LIKPIALNII-----REKVNKC 1604
Db 1963 LDQQLDQLFHWVNVAGCVGWGEVQEQEEAVHKC 1997

RESULT 35
Tl3030
microtubule binding protein D-CLIP-190 - fruit fly (Drosophila melanogaster)
C/Species: Drosophila melanogaster
C/Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
C/Accession: Tl3030
R/Lantz, V.A.; Miller, K.G.
J. Cell Biol. 140, 897-910, 1998
A/Title: A class VI unconventional myosin is associated with a homologue of a microtubul
A/Reference number: Z17588; MUID:98139549; PMID:9472041
A/Accession: Tl3030
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 1-1690 <LAN>
A/Cross-references: EMBL:AF041382; NID:g2773362; PID:g2773363; PIDN:AAB96783.1
A/Experimental source: strain Oregon R
C/Genetics:
A/Cross-references: FlyBase:FBgn020503
C/Keywords: cytoskeleton

Query Match 3.2%; Score 274; DB 2; Length 1690;
Best Local Similarity 18.7%; Pred. NO. 0.00024;
Matches 291; Conservative 278; Mismatches 562; Indels 422; Gaps 70;
Qy 372 DKLFLKQHCEPQEGV-IKIKASSLSTYKIAEQDFSYFFDDPPTFFSPANRRRGRPP 430
Db 286 DDIRYF---ECKPKGVFPIAKVSL-----SPSS----- 312
Qy 431 KRIHSQEDNVANKOTLASYSKATKRDKL-LKQEMWKS-----LAFKAK-----L 477
Db 313 KKTLSRTGSRSLTSIGTMSNIATTSRMRMAQQRKSSTPVKPILATPKSQFSMDL 372
Qy 478 KREKADALE---AKKXKEDKEKKRELKIVEERLKKKEERLKVREKERE----- 529
Db 373 LREKQHVLEKLMVERDLREDQAQNALQLOKNINELKARIVELESALGNERRKTEELQCS 432
Qy 530 -----KLREERKRVLEVILKQSKPREDMCEDDLKELPEPTP-VKTRLPPEIFGDAL 579
Db 433 IDEAQFCGDELNAQSOVYKEI-----HDL-SKITKLVSATPSLOSILPPDLPSDDG 484
Qy 580 MVLEFNAPGELFDLQDFPQGVTLVELEALVGNDSGPICELLFFFLTAIFAIAEE- 638
Db 485 ALQEBIAQLEQKWTIQQKEVESRTAEQLEEE-----QRLRENV 522
Qy 639 ---BEEVAKEQLTDADTKGCSLSKSLDLSCTLSEILR-LHILASGADVTSANAKRYQKR 694
Db 523 KYLNEQIATLQ-SELVSKDEALEKPSLSECGIENLRRELLALKEENKQAEQABFTRK 581
Qy 695 GGFOATDACCMELRSLNSPLVKLSSTSVYDLTPGKMKILHALCGKLLTLVSTRD---- 750

Db 582 ----LAEKSVLEVRLS--SELQNLKATS--DSLSESRVNSKDE-CEILQTEVRMRDEQIR 632
Qy 751 -----FIEDYVDILRQAQOEREL--KAEQHRKREEEAARLRK 787
Db 633 ELNQLDEVTTQLNVQKADSSALDMLRLQKEGTEEKSTLEKTEKELVQIKEQAATLQ 692
Qy 788 RKEKLEKEQEKME--KQEKLEDEORNSTADISIGEEEREDFDTSI-----ESKOTEQ 840
Db 693 DXEQ-LKQISDLQKLAQOEKLVREKTENAINQOL-----EKESIEQQLAKONELEDQK 748
Qy 841 KELQDDMTFEDDDPGSHKRRGRGQNGKFEFTRQEQINCVTRELTADAEEALKQEH 900
Db 749 KQSESEVHLQBIKAQNTQKOLELVESGES-LKLQOQOLEEKTGLGHEKLAALAEELKKE 807
Qy 901 ---ORKEKEL-----LEKTOAIACTNIPLGRDRMYRYWIFPS 937
Db 808 TIIEKEQEQLOQLSKAESALKVVQVQLEQLQQAAS-----GEGSKTVAKLHDE 862
Qy 938 IPGL--FIEDYSGLTEDMLLPSPSSFQNNVQSQDPQVSTKTGPPLMSESTNIDQGRD 995
Db 863 ISQLKSAEETQSEL-----KSTESNLEAKSKOLEAANGS--LEEAKKSGQ-LQE 910
Qy 996 HSQLPXPHKPNR--WCFYSSCEQLDQLIBALN-----SRGHRESALKETLQEKSR 1047
Db 911 QITKLESEVEETOALSSYHTDVESKTKQLEAANAALAEKVNKVAESRAEASDLQDKVE 970
Qy 1048 CA-----QLARSEBEKFH-----FSDKPODPSKPTYSRGRSSNAYDPSOMCAEKQLEL 1095
Db 971 ITDTLHAELOAQRSSSALHTKLKSFDEIATGHKELTSK---ADAMSQEWLQKEKEVQ- 1026
Qy 1096 RLRFPLDIEDRIYQGTGLGAIKVTRHWSALSGRYELLSEENKENGIIKTVNEV-- 1153
Db 1027 ELRQLOQDSQDS--QTKLKA-----EGERKEKSFES-----IKNQEEVTK 1066
Qy 1154 ---EEMEIDEQTVIVKD-----RLLGKITETPTSTVSTNA- 1185
Db 1067 AKTENLELSTGTQTTINDQLERLEITNAELQHEKEMASEDAQKIADLTKLVEAIQVAN 1126
Qy 1186 --STPQSVSVVHYL-----AMALFQIE-----QGIERRFLKAPLDASD 1222
Db 1127 ISATNAELSTVLEVLQAKSETNHIPELFEMADMNSERLIEKVGTGIEELKETHLQDE 1186
Qy 1223 SGRSYKTVLDRW-----ESLSSASLSOVFLHLSTLDRSVTWSKSL-NARCK 1270
Db 1187 ROKFEELEEKLAQAOQSEQKLQOESQTSKEKLTETQOSLOELQDSVKQKELVQNLEK 1246
Qy 1271 ICRKKG--DAENMVLC-DGDRGHHTYCVRPKLKTVPEGDWFCPECRPKQRRLSFRQ 1327
Db 1247 VRESSIIIEAQTKLINESNVQLENKTSCLKETQDQLLESQ--KKEKQLQEEAAKLSGELQ 1304
Qy 1328 PSLESDVEDSMGGEDEVDGDEEGQSEEEVEVEQ-----DEDDSQE 1372
Db 1305 QVQEANGDIKDSLKVKEVLVLEKLAQTSQDAQOATNKLQELQELVLVQSENEGNOG 1364
Qy 1373 B-----EVSUPKGRPOVRLPV-----KTRGKLSSESFS-SRGQOQEPGRYPSRQ 1418
Db 1365 ESLAVTEKLOLEQANGLKELCKQNGKELQKLOKLDSENVLESQKSHNEIQDKLEQ 1424
Qy 1419 STPKTTV-----SSKTGRSLRKINSAPPTETKSLRIASRSTRSHGHPLOADVVELLSPR 1474
Db 1425 AQQKERTLQEBTSKLAEQLSKQANBELQSLQ-----QKLLLE----- 1465
Qy 1475 KRRGRKANNTPENSPNPNFRVVIATKSS-----EQ-----SRSVNIASKLSLOES 1520
Db 1466 --KGNFDTQLAEYQKVIDEMDDAASVKASALLEQLQNRVAELETALROANDQKTYLET 1523
Qy 1521 ESKRCRKRQSPPEPVTLGRR-----SSGRQGVHLSAFE-----OLVVLE 1563
Db 1524 KELRRQLESLELESREVLSLKAQMGNGASSRSGKDEVEDIETSLAKINFLNSIADM 1583
Qy 1564 VRHDDSWPFLKLVSKIQVDPYDI-IKPIALNIIREKVNKVEYKLAASEFIDDIELMPSN 1622
Db 1584 QOKNDA-----LKAKVQTLPLPMTKPHAFDALTKR-----KPAPRLFCDI----- 1626

Qy 1623 CFENPNRNTSBAKATRLQAFPHIQAKLGLHVTFS-----NVDQVSTPPAAKK 1671
Db |:::|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
1627 CDFDOHDTEDCP-----IQSGEDQDYSPSSSNNEKERKLPA PRK 1669

RESULT 36
JC5837

364K Golgi complex-associated protein - rat
C:Species: Rattus norvegicus (Norway rat)
C>Date: 05-Mar-1998 #sequence_revision 13-Mar-1998 #text_change 09-Jul-2004
C:Accession: JG5837
R:Toki, C.; Fujiwara, T.; Sohda, M.; Hong, H.S.; Misumi, Y.; Ikehara, Y.
Cell Struct. Funct. 22, 565-577, 1997
A>Title: Identification and characterization of rat 364-kDa Golgi-associated protein re
A:Reference number: JC5837; MUID:98093490; PMID:9431462
A:Accession: JG5837
A>Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-3187 <OK>
A:Cross-references: UNIPROT-Q63714; DBJ:D25543; NID:9516825; PID:NBAO5026.1; FID:95168
C:Comment: This protein plays a role in the formation and maintenance of the character
C:Superfamily: giantin
F:49-549,624-1176,1238-1707,1763-3114/Domain: coiled-coil leucine zipper #status predic
F:3165-3187/Domain: membrane anchor #status predicted <MAD>

Query Match 3.2%; Score 274; DB 2; Length 3187;
Best Local Similarity 20.1%; Pred. No. 0.0005;
Matches 315; Conservative 259; Mismatches 618; Indels 376; Gaps 72;

Qy 213 TYQEALSEKKARQNLOSFPEPLIIPVLVLTSLTHRSRLHEICDDIFAYVKYDRFYVEETV 272
Db |:::|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
1361 TVQTMEBEQELIKALHT-----QLEMQAKEHEERLKQVVEICELKKQKEUEEES 1412

Qy 273 EVIRNNGARLCQTTILEVLPSSHQNGFANGHVNSV-DGETIIISDSDSETQSCSFNQKK 331
Db |:::|:|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:
1413 KAKQLQRKLQAALSRKEALKENKSLOEQLSARDADAVEHLTKSLADVESQ-VSVNQOEK 1471

Qy 332 KDATDPILLFKYKVQPTKELHESAIVKATO--ISRRKHILFSRKCLKFLKQHCEPOGEVI 389
Db |:::|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:
1472 ----DALLGLKAL---LQERDKLIVEMDKSLLENQSLGGSCSLKLAGLGTDEKELM 1524

Qy 390 KIKASSLSTVKIAE---QDFSIFYFPDPTTFITSPANRRRGPRPKRIHIISOQDNVANQK 445
Db |:::|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:
1525 K-ELESVRCSKIABSTWQE-----KHKELQKEVEV---- 1554

Qy 446 TLASYRSKATKERDKLLAQEBMKSLAFEK---AKLRKADALEAKKKEKEDKEKKREE 501
Db |:::~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:
1555 LLQSYEN-VSNAERI-KHVVESVRQEKQEVYAKLRSAESDKRE-REKQLQDAEQEMEE 1610

Qy 502 LK-----KLIVERELKKKEERLKVRE-----KEREKUREKKYVEY 541
Db |:::~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:
1611 MKEKORKFAKSKQOKILLEENDRLRAEAQPVGANMESALLSSNASLUKSELERITLE 1670

Qy 542 LKOWSKPRE-----DMECDDLKELPEPTPVKTRLPPIFGDALMWLE 583
Db |:::~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:
1671 YKTLKSEFEALMAEKNTLSBETRNLKQVDAQELKQASLETTEKSDPKDV-----IEE 1724

Qy 584 FLNAGBELFDLQDFPDGVTVLEVLEALVGNDSGPCLCE-----LFFFLTAFQ---A 634
Db |:::~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:
1725 VTEAVGVGSQBQDSLSENAKLEDAAETLLANSAPGVSETFSSHDDINNYLQQLDLQGR 1784

Qy 635 IAEIEEEVAKEQ-----LTDATKGCSLKSOLDSCITLSEILRLHLIASGAD 681
Db |:::~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:
1785 IAELEMESKQDRSLSQTLENEKNALLTQISAKOSELKLE-----EEVAKINML--NOQ 1836

Qy 682 VTSANAKRYQKRGGFDATDACMELSLNPSLVKKLSSVSVY--DLTPCKMKILHALC 739
Db |:::~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:
1837 IQEELSRVTKETAEEKDD--LEERLMN-QLAELNGSIGNYQDVTDQA----- 1884

Qy 740 GKLLTVLSTRDFIBDYVDILRQAKQEFRELKAQHRKEREEAAAIRINKRKEKL----- 793
Db |:::~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:~|:
1885 -----IKNEOLESEMONLRKCYSLEEEBK-OOLVKEKTQVSEIRKEYMEKIOGAOK 1936

Qy	794	-----KEQEKWKKEQEKLEDE-----QRNSTADISIGEEERDFDT--	831
Db	1937	PGSKIHAKELOELLKEQOEQVKLOKDCIRYLGRISALEKTVKALEFVHTSESQKLDATK	1996
Qy	832	-----SIESKDTBEKELDQ--DMTEDEDDPGSH-----KGRGRKGQNGFKFEFTROEQ	879
Db	1997	GNLAQAVEHHKKAQAEUSSFKILLDDTQSBAARVJADNLNKKLEQSNKESIQSIIKOKD	2056
Qy	880	INCVTRELLTADBEELAKQEHORKEKELLEKIOSAIACTINPFLGRDMRYYRWFPSIP	939
Db	2057	EDLLR-----LEQAEKHKEKQWQEKJDA-----LHREKAH-----	2090
Qy	940	GLFIEEDYSGLTEDMLLPRESSFONNVQSDPOVSTKTGEBPLMSESTSNIDQGRDHVSQ	999
Db	2091	-----VEDTLAEIQVSLTRKDKMKELQQSLD-----STILAQAAFTKSMSSL-QQDRD----	2138
Qy	1000	LPXPVHKPNWCYSSCEQLDOLLEALNSRGHRESALKE---TLQEKSEICQALARFSE	1056
Db	2139	--RVIDEAKKW-----EQ--RFGDAIQTK--EEVRLKEENCFTALKDQLR---OMTIHME	2184
Qy	1057	E-KFHFS-----DKFPQDSKPTYSGRSSNAYDPSQCAEQLELRDLDFLLDIEDRIYQG	1111
Db	2185	ELKITVSRLEHDEKIEWESKAQTELHQOQKAYDKLQ---EENKELMSQ---LEAQGLYHD	2238
Qy	1112	TLGAIKVTDRI-----WRSALSGRYELLSEENKENGIIKTVNEDV-----BE	1155
Db	2239	SKNELTKLESELKSLKDQSTDKNLSLEKCR-----EHENNLGIIIKQOEADIQNCKFNCEQ	2294
Qy	1156	MEID-----EQT-----KVIVKDR-----LLGIKTETPTSTVSTWASTPOSVSSVVHYLAMA	1201
Db	2295	LETDLTASRELTTRLHDEINVKQKIIISLSGKEEAIQVAI---AELHQOQSHKSEIKELN	2352
Qy	1202	LFQTEQ--GIERRFLKAPLDASDGRSYKTVLDWRWRESLSSASLSQVFLHLSTL----	1254
Db	2353	LSQEEBENLTLEENKGAVEKTKLTLEATETIK---KESLEQKALDLSFVKSMSSLODDR	2409
Qy	1255	DRSVIWSKSTLNARCKIRKGD-----AENMVLDGCD--RGHTTYCVRPKLKTVPEBG	1306
Db	2410	DRIVSDYRQLEERHLSVILEKDELIQDAEAENNKJAKEIRGLRGHMDLNSENAKLDAEL	2469
Qy	1307	DMFCPECRKQRCRLSFQO--RPSLESDEDEDSMGCEDEVDGDEEGEESGSEEEVEE	1364
Db	2470	IQYRDDNEVITTKDSQORQLLEAQLOQNKELRNECVKLEGRUKGSAEKOSIQMSLDAL	2529
Qy	1365	QDEDD-----SQEEEVESLPKRG-----RPOVRLPVKTRGKLSSSFSS---RGQOQEP	1409
Db	2530	QENQGLSKKEIKFKQLTALHEGALVAVYHAQLRVREEEVQKLTAAALSSQKXTVDLOE	2589
Qy	1410	GRYPSRSQOSTPKTTVSSKTRSLRKIN-----SAPPTETKSLRIASRSTRHSHGPIQAD	1464
Db	2590	ELVCVQKEASKYSEIEDKLKRLKHLHHNAGIMRNRETETAERVAE-----LARD	2640
Qy	1465	VFVELLSPREKRGKRKANNTTPNSPNFPNFRVIATKSSQESQSVNIATSKLSLOESSEKR	1524
Db	2641	L-VEM-----EQKLLTVTKENKOLTAQIOAFGKSMSSLODSRDHATE-ELUSDUKKY	2690
Qy	1525	RCRKROSPSPVTLRRSGROGGVHELSAF-----EQJLVVELVRHDDSWPF	1572
Db	2691	DASLKELAQ--LKGRODLGRESVDLSQAFFLPTTSENISSRLEKLNQOLISKDEQ--L	2744
Qy	1573	LKLVSKIQ	1580
Db	2745	LHLSSELE	2752

RESULT 37

T16416

hypothetical protein F52B10.1 - *Caenorhabditis elegans*

C;Species: Caenorhabditis elegans

C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 09-Jul-2004

C;Accession: T16416

R;Wilcox, L.

Db 932 TSGDPKK-----SAPSTPVVRDAGSDLVQAQMSNQ-----PNLGLRKL 969
Qy 429 PPKRIHISQEDNVANKQTILASYSK-----ATKRDKLLKQBEKMSLAFKAKLREKAD 483
Db 970 PRIKKSSALQIQNHQPHSNAQSTPSTSTHQAMPKDKKKKKKKKEKEEREREA 1029
Qy 484 ALEAKKKKKEDKKREELK--KIYEEBLKKKKERLKVREKEREKLR-----EKKK 537
Db 1030 RREMKRKETKEBRNKKEMERAKRLDEQERKKERKKERDERKKERKVRKAKKEK 1089
Qy 538 YVEYLKQSKPRMEDCECDLKEPPTPVKTRLPPEIFGDALMWLEFLNAGELFDLQDE 597
Db 1090 KKKHRKGGSSDSDSDSDELDELVRKSTKE----- 1120
Qy 598 PDGVTLEVLVLEALVGNSEGLCELLFFFLTAIFAQIAEEEEEVAKQ-----LTDADT 652
Db 1121 ----MTQBEKDHLALLSKGGIENLKSRRSDKRAHDSPEKMQQKQSOORVLISSDD 1176
Qy 653 KCSLKSLLDSCITSEILRLHLILASGADVTSANAKRYQKGGFDATDDACMELRLSNP 712
Db 1177 EG-----GKGGDKGNS-----SNGESDSEKADLPPPPAPP 1207
Qy 713 SLVKKLSSTSVYDLTPGKMKILHALCGKLLTLVSTDFIEDYVDILRQAKQERELKAE 772
Db 1208 SLSE-----SADQLKVLKEREKELTSSDDEHDNDAGIHOORLTEDRENKR 1257
Qy 773 Q-----HRKEREAAAIRKKEE-----KLKEQKQKKEKQKLEKDEORNSTADI 819
Db 1258 QKSLTAYSSDEQERKNVPKMRDDSEDAAKHPGWSAKODQKQKLEHRRSS----- 1313
Qy 820 SIGEEREDFTSISKQTEQKEL-----DQMFTDEDDPGSHKGRKRGKQNGKFKEFT 875
Db 1314 ----ED-----ESKNAKRDPRDIPHEVDSDEETEDGSRSR----- 1346
Qy 876 ROEQINCVTRELTLADEEALKQERKEKELLEKIQSALACTNFIPLGRDRMYRYWIF 935
Db 1347 KQSTSSITSN--VTAK-----RKEK----- 1365
Qy 936 PSIFGLFIEDYSGLTEDMLPRPSSFQNVQSDPQVSTKTGPELMSESTNIDQGRD 995
Db 1366 -----SGKTPLRIVPEP-----TGTPLLSP----- 1385
Qy 996 HSVQLPKVPKPNRWCFSCEQLDQLLEALNSGRHSALK--ETLQEKSR-----ICA 1049
Db 1386 ----KILSPK-----LSPKSTSTSKRSSIDHENLISPRQRNRTTSST 1426
Qy 1050 QLARESEKHFSDKQPDSPKTSYGRSSNAYDPS-----QMCQKQLELRDLFL 1102
Db 1427 STATTSSKHEALSIPEKPLSPVTAKSSVSSIDDPISIRDEFMSNSAADSPMTTGRPMVL 1486
Qy 1103 DIEDRIYQGTILGAIKVTD-----RHIVRSAL-----ESGRYELLSEENKENG 1145
Db 1487 -----TKAAMKAFNSTPPKVSYSLLIDCYMLGMWAKNSSSGQHDSSGSSDSSS 1537
Qy 1146 I--KTVNEDVEMEIDEQTVIVKDRLLGKITETPTSTVSTNASTPQSVSSVVHYLAMALF 1203
Db 1538 SDGTSSTSDSDSDDEVPKQTEPV-----TSIPVWASDNGS--PENV----- 1575
Qy 1204 QIEQGIERRFLKAPLDASDSRSYK--TVLDRWRSELLSSASLSQVFLHLTLDRSVIWS 1261
Db 1576 -----VVETPSIVSQTPREPFTTISEQSSS----- 1602
Qy 1262 KSLINARCKICRKGDAENMVLCDGCRGHHTYCVRPKLKTVPEGDWFCEPCRPKQRCR 1321
Db 1603 -----EPEAVPE-----CPASVE----- 1616
Qy 1322 LSPRQPSLESDVED--SMGGEDEVGDDEE-----EGQSEEEYE-----V 1363
Db 1617 ----PQMTSONVPEVPESEHDSHGDSVAVESQQQPLEHQEKEELEENKILDVAA 1670
Qy 1364 EQDEDDSEEE--EVSLLP--KGRPQVRLPVKTRGKLSS-----SFSRGGQQQEPGR 1411
Db 1671 EHHEEQVGDESDSVESIPAPSDPDPVTOAQEKSAHTLISDQETDQAVQSFIDEEADE 1730

Qy 1412 YPSRSQ--QSTPKTTVSSKTRSLRKINSAPPTETKSLRIASRSTRSHGPIQADVFVEL 1469
Db 1731 FPQYPDFGISTNEKSVSGDPHNIK-----PTE-----PLNNGHTDL-----L 1768
Qy 1470 LSP-----RRKRGKRSANPTPENS-----PNFPNFRVIATKSSEQRSRVNIASKLSIQ 1518
Db 1769 FSPSSAHASQSTSEDDMEEDSELVVMKEVMEQVIA-----QEVHVPSESPM 1821
Qy 1519 ESESK-----RRCRKROSP--PSPVTLGRSSRGQG-----VHELSAFQLVVE 1562
Db 1822 EEEVKLETSPVKKEEPIKMEESPQOTPTDILSNNESSQDTPGAVNNHLHENHDAVOTPIQ 1881
Qy 1563 L---VRHDDSWPFLK-----LVSKIQVPDYDIKPIALNIIRKVNKCEY 1606
Db 1882 LQASQHQVAQSPRPAPVAPDSQQNGPVLVSQSQSPSPMSQSQSDMAQNLI----- 1932
Qy 1607 KLASFEIDDIELMFSNCFEYNPNTSEAKAGTRLOAFFF--IOAQKGLGHLVTP 1657
Db 1933 -LSSKDINDLAALKLHK-----NPEALAAQATRGDCSGIFQHLHLHAQNGQNMT 1980

RESULT 39

A47297

myosin heavy chain form B, nonmuscle - African clawed frog

C:Species: Xenopus laevis (African clawed frog)

C>Date: 22-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 09-Jul-2004

C:Accession: A47297; A55441

R:Bhatia-Dey, N.; Adelstein, R.S.; Dawid, I.B.

Proc. Natl. Acad. Sci. U.S.A. 90, 2856-2859, 1993

A>Title: Cloning of the cDNA encoding a myosin heavy chain B isoform of Xenopus nonmuscle

A:Reference number: A47297; MUID:93219383; PMID:8464900

A:Accession: A47297

A>Status: preliminary; not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 1-1992 <BHA>

A:Cross-references: UNIPROT.Q04834; GB:L09740; NID:9214623; PIDN:AAA49915.1; PID:G214624

A:Experimental source: XTC cells

A>Note: sequence extracted from NCBI backbone (NCBI:128722)

J,Kelley, C.A.; Oberman, F.; Yisraeli, J.K.; Adelstein, R.S.

R. Biol. Chem. 270, 1395-1401, 1995

A>Title: A Xenopus nonmuscle myosin heavy chain isoform is phosphorylated by cyclin-p34

A:Reference number: A55441; MUID:95138137; PMID:7836406

A:Accession: A55441

A>Status: preliminary

A:Molecule type: protein

A:Residues: 198-232 <KEL>

C:Superfamily: myosin heavy chain; myosin motor domain homology

C:Keywords: nucleotide binding; P-loop

F:88-787/Domain: myosin motor domain homology <MWOT>

F:178-185/Region: nucleotide-binding motif A (P-loop)

Query Match 3.2%; Score 272; DB 2; Length 1992;

Best Local Similarity 20.0%; Pred. No. 0.00036;

Matches 266; Conservative 210; Mismatches 483; Indels 368; Gaps 56;

Qy 361 QISRR-KHLFRDRLKFLKQHCPEQGV---KIKASSLSTYK--IAEQDPSYFPDPP 414

Db 862 QVTQEEELVADBELKVKREKQKVGELVDMQKQQLVEEKNIILAEQLHA----- 914

Qy 415 PTFIFSPANRRRPPKRIHISQEDNVANKQTLASYSRKATKRDKLLKQBEKMSLA--- 471

Db 915 ETELFASAEEMRA---RLAIKKQEMEELRDL--EIRMEEEERNQVILQNEKKKQWTHVQ 969

Qy 472 -FEKAKLREKADALEAKKKEKEDKREBELKIVIEERLKKKEERLKVREKEREK 530

Db 970 DLEEQDDEEAAQKLEKVTAEAKIKKMEEDILVLEDQNSKFLKELKLEERIAESTSQ 1029

Qy 531 LREKRYKVEYLKQMSKPRDEMCDDKELPEPTPVKTRLPPEIFGDALMWLEFLNAGE 590

Db 1030 LAEEEEKAKNIK--LNNKQEMMTSDLEERLUKKEE--KTRQELE---KAKRKLD-----GE 1078

Qy 591 LFDLQDEFPD-GVTLEVLLEALVGNDSGELCELLFFFLTAIFAIAEBEEBEVAKELTD 649

```
Db 1079 TTDFQDQIAELQAOQIEELKQLAKKEE-----LQAALARGDEVILQ----- 1120
Qy 650 ADTKGCSKSLDLOSCITSEILRLHILASGADVTSANA--KYRYKRGFGDATTDACME 706
Db 1121 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 707 LRSLNPVLVKLSSTSVYDLTPGKMKILHALCGKLLTLVSTRDFIEDYVDILRQAKQEF 766
Db 1163 LEALKTELEDITDYTA-----AQEEL 1183
Qy 767 RELKAEQHRKEREAAAIRKRKEKLKEQKMKKEKLEKEDQKRNSTA--DISTGEE 824
Db 1184 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 825 EREDFDTSIE---SKDTEQKELQDMFTEDDPGSHKRGKRGKRGQNGKFKETRQEQI 880
Db 1229 QAKRFKNLEKNKQSLDNKELATEV-----KSLQQWKAESYKRRKLEGGV 1276
Qy 881 NCVTRELLTADBEALQEHQKQKELLEKIOSAI--ACTNIFPLGRDMRYRYWIFPSI 938
Db 1277 QELHAKVLEGRLRADVWEKSKLQNELNVSSILLEAEKKGIKLAKDVAMESQL-QDT 1335
Qy 939 PGLFIEDYSGLTEDMLPRESSFQNNVQSQ----- 969
Db 1336 QELLQETETKLNQSSRIQLEEEKNVLEQEQEEBEAEKSLKQILSLQSLQTEAKKV 1395
Qy 970 DPQVSTKGT-----EPIMS-----ESTSN-----IDQGRD 995
Db 1396 DDEVGTIEGLEBEVKKLKDKTEGLQRLKEKIIAYEKLKTKNRLQOBLDLMVD---LD 1452
Qy 996 HSVOLPKVPVHPNRCFYSSCEQDLQI---EALNSRGHRESALKETILLOKSRIQAQLA 1052
Db 1453 HQQIVSNLEKKQK-----KFDQLLAENKTSARHAERDRAEADAREKETKALSILA 1504
Qy 1053 RFSEKHFHSDKPOPSDKPTYSR-----GRSSNAYDPQMCABKQLELRDRDL 1101
Db 1505 RALDEALEAQDEFERLNLQLAEMEDLMSSKDDGVKNVHELEKSKRALDQQVE-EMRTQL 1563
Qy 1102 LDIEDRIYQITGLGAIKVTDRIHWSALESGRYELLE-----ENKENGIIKTV 1149
Db 1564 EELEDEL-QGTEDA-----KLRLENNQAMKAQFERDLQTRDEQNEKKKALVQV 1613
Qy 1150 NEDVEEMIDROTQKVIIVKDRLLGKITETPSTVSTINASTPQSVSSVVHVALMALFQIEQI 1209
Db 1614 RELAELE-DER-----KQRAVAV-----AIKKKLEMDMKDFESQI 1648
Qy 1210 ERRELKAPLDASDSGRSYKTVLDRWRRESLLSASLSQVFLHSLTLDRSVIMWSILNARC 1269
Db 1649 E-----AANKGR--EDAIKQLRK-LQATKDYQRELEEARASRDDIFAQSKENEK- 1695
Qy 1270 KICRKGK-DAENMVLCDGCRGHITYCVRPKLKTVPEGDWFCPCGRPKQRCRRLSFRQRP 1328
Db 1696 ---KLKGLEAETLQLE-----ELASERSRRRAEQRD 1726
Qy 1329 SLESDEVEDSNGCEDDDEVDGEBEGQSEEEYEVEQDDSDQSEEEVSLPKRGQPURL 1388
Db 1727 ELA--DEISNTSGSKALD---EKRRLEARIAHLEEBEESQNMEL-LNDRFR-KTTL 1779
Qy 1389 PVKTRGKLSSFSRSGQOQEGRPVPSRQQSTPKTTVSSKTRSLRKINSAPPT----- 1442
Db 1780 QVDTLNSELAAERSSGQSENAQQLERQNKELKAKLQELSGSVKSKFKATITLESKIA 1839
Qy 1443 -----ETKSLRIASRSTHSHGFLQADVVELLSPPR-----KRRGRKSANTPENS 1489
Db 1840 QLEEQLEQAEKERVASNLVARTKELK-K-EVFMQVEDERRHADQYKEQMEKANTRMKOLK 1898
Qy 1490 ENFNFRVIATKSSFSQR-----SYNVIASKLSLQSESKRCKRKQSPSPVTLGR 1541
Db 1899 RQLEAEAEATRANASAKLORELDADATEANEVLSREVSTLKNLRR-----GGFVSPS- 1952
Qy 1542 RSSGRQG 1548
||| |||
```

```
Db 1953 SSSRSRG 1959
RESULT 40
S49883
nuclear protein STH1 - yeast (Saccharomyces cerevisiae)
N;Alternate names: cell division control protein NPS1; protein YI126.03; protein YI126.03
C;Species: Saccharomyces cerevisiae
C;Date: 28-May-1993 #sequence revision 24-Feb-1995 #text_change 16-Aug-2004
C;Accession: S49883; S28406; S22777; S38377
R;Hamlyn, N.; Churcher, C.
submitted to the EMBL Data Library, November 1994
A;Reference number: S49881
A;Accession: S49883
A;Molecule type: DNA
A;Residues: 1-1359 <HAM>
A;Cross-references: UNIPROT:P32597; GB:Z47047; EMBL:Z46833; NID:g603997; PID:g7633220; MI
R;Tsuchiya, E.; Uno, M.; Kiguchi, A.; Masuoka, K.; Kanemori, Y.; Okabe, S.; Mikayawa, T.
EMBO J. 11, 4017-4026, 1992
A;Title: The Saccharomyces cerevisiae NPS1 gene, a novel CDC gene which encodes a 160 kD
A;Reference number: S28406; MUID:93010997; PMID:1396591
A;Accession: S28406
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-1359 <TSU>
A;Cross-references: EMBL:D10595; NID:g218448; PIDN:BA001446.1; PID:d1001921; PID:g218449
R;Laurent, B.C.; Yang, X.; Carlson, M.
Mol. Cell. Biol. 12, 1893-1902, 1992
A;Title: An essential Saccharomyces cerevisiae gene homologous to SNF2 encodes a helicase
A;Reference number: S22777; MUID:92195336; PMID:1549132
A;Accession: S22777
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-545,'T',547-1009,'R',650-1009,'R',1011-1072,1080-1359 <LAU>
A;Cross-references: EMBL:M83755
R;Laurent, B.C.; Yang, X.; Carlson, M.
submitted to the EMBL Data Library, February 1992
A;Reference number: S38377
A;Accession: S38377
A;Molecule type: DNA
A;Residues: 1-545,'T',547-1009,'R',1011-1072,1080-1359 <LA2>
A;Cross-references: EMBL:M83755; NID:g172763; PIDN:AAA35120.1; PID:g172764
C;Genetics:
A;Gene: SGD:STH1; NPS1
A;Cross-references: SGD:S0001388; MIPS:YIL126w
A;Map position: 9L
C;Superfamily: bromodomain homology
F;1278-1333/Domain: bromodomain homology <BRO>
Query Match 3.1%; Score 271; DB 2; Length 1359;
Best Local Similarity 18.8%; Pred. No. 0.00025;
Matches 298; Conservative 243; Mismatches 536; Indels 448; Gaps 73;
Qy 270 ETVIEVIRNNGARLOCTILEVLPPSHQNGFANGHVNSVDGETIIISDDSDSTQSCSFQNG 329
Db 7 ELMTVANNPTTTVAALAAVAASSETNG-----KLGSERQPSITTPK 49
Qy 330 KKDAIDPLFKYKV---OPTKELHESAIVKA-TQISRRKHLFSR--DKLKLFLKQHCE 383
Db 50 RSSAQLEQLLYRYRAIQNHKPKENKLEIKALIEDTFRNSRDQDIYETKLTDLRKSIDKGFQ 109
Qy 384 PQEVIKIKASSISTYKIAEQDFSYFFPPDDPPTTFSPANRRRG-----RPPKRIHI 435
Db 110 YDEDLN---KHLVALQLLEKD-----TDVPDYFLDLPDTKNDNTTAIEVDYSEKKPKI 161
Qy 436 SQEDNVANKQT-LASYRSKATK-----ERDKL----- 461
Db 162 SADFNAAKSLGLSKESKFNATKALTALGDPDTEIRISARISNRINELERLPANLGTYSLDCC 221
Qy 462 ---LKQEMKSL--AFKAKLKREKADALEAKKEKEDK----- 495
Db 222 LEFITKODLSRMDTFKIKALVELKSLKLTQKSIKQLINNVAQAHNIPYLRDSPF 281
```

Qy	496	----	EKKBEELKKIV--BEERLKKKEERLKVREKEREKRLREBKRYVYLK-----Q	544
Db	282	TAAARSQVIRSKVIVPQTVLAELEERQQLLEKKERKLNHLQKINSIIDFIKERQSEQ	341	
Qy	545	WSKP-----REDMECDLKELPPTPVKTLRPEIFGDALMVLEFN-----	586	
Db	342	WSRQERCFOFQELGASLHNQEKEDQKRIER--TAKORLAALKNDDEAYLKLJQYKDT	399	
Qy	587	AFGELFDLQDBFPDGVTLVELEAALVGNDSGPCLCELLFFLFTLTAIFAIAEBEE-	640	
Db	400	RITQLLRQTNSEF-----LDSLSEAVRAQQNEAKILH-----GEEVQPIITDEEREKTDY	448	
Qy	641	EVA---KEQLTDAOT--KGCSLKSLDLSDCTLSILRLH-----ILAS-----GADVTSA	685	
Db	449	EVAHRIKEKIDQPSILVGGTLEKYQLRG--LEWVMSLYNNHNLGILADEMGLGKTIQSI	506	
Qy	686	N-AKRYQKR--GGF-----DATDOACMELRLSNPSLVKKLSTSVYDLTPCEKMKIL	735	
Db	507	SLIYTLFEVKDIOGFVLVPLSTITNTWLEFEKWAPSL-----NTIYKGTPNQRHSLQ	563	
Qy	736	HAL--CGKLLTLVSTRDFIEDYVDILROAKQBFRLKAKQHRKEBEREAAAARIKRKEEKLK	794	
Db	562	HQIRVGDFDVLTTIYIYIKDKSL--SKHDWAHM-----II	596	
Qy	795	EOEQKMEKQEKLK--EDEORNSTADISIGEEREDFD-----TSIESKDTQEKQLDQ	845	
Db	597	DEGRHMAQSKLSFTISHYVYRTRNRLITGTPLQNNLPALWALLFVLPKIFNSAKTFE	656	
Qy	846	DMFTEDEDDPGSHKGRGRKGQNGFKFEFTQEQINCVTRELLTADDEEALKQEHQREK	905	
Db	657	DMFWNTPANTQBEK-----LELTEEETL-LIIRLKHVLPFRLLRLRLKCEVEK	704	
Qy	906	ELLEKIOSAIACTINIFPLGRDRMYRRYWIPPSIPLGFIEDYSGLTEDMLLPSPSSFONN	965	
Db	705	DLPDKVEKVIK--KLSGL--QOOLYQO--MLKHNLVFGAGTEATKGGI-----KGLNNK	755	
Qy	966	VQSDOPQVSTKGTGPELMSSESTNIDQGRDRHSVOLPKPVHKPNRWCIFYSCQELDQLEA	1025	
Db	756	IM-----QLRKICNHPFVDFEVEGVVNPSPRGNSDLLFRVAGK-----PELLDRVLPK	802	
Qy	1026	LNSGHRSALEKFTLLQEKSRICAQLARFSEKHFHSDKPDQPSKPYISGRSSNAY---	1082	
Db	803	FKASGHR--VLMFPQMTQVMDIMEDFIRMOLKTYRLD-----GSTKTEERTEMLNAFNAP	856	
Qy	1083	DPSQMAEKQLELRLDRFDLLIEDRIYQGTUG-----AIKVTRHWRSALESGRY	1133	
Db	857	DSDYFC-----FLISTR-----AGGLGLNLQTAADVIIIFDTWNPHQDLQAQDRA	901	
Qy	1134	ELISEENKENGIIKTVNED--VEEM-----EIDEQTKVIVKORLLGKITETPTSTVSTNA	1185	
Db	902	HRIGOKN-EVRIILRLITTDSVEEVILERAMQKLDIDGKVIQAGKF-----DNK	948	
Qy	1186	STPOSVSVVHYLAMALFOETGIERFLKPLDASDSGRSYKTVLDRWRESLSSASLS	1245	
Db	949	STABE-----QEAFLRLISETNRDRDDDKA-ELDDDELNDTLTARS---989		
Qy	1246	QVFLHLSTLDRSVIWSKILNARCKICRKKGAENMVLCDGCRGHHTYCVRPKLKTVPE	1305	
Db	990	-----DEKILFDKI--DKERMQERADAQAQGL-----RVPPPKLIQDE	1027	
Qy	1306	GDWFCPCRPQRCRLRSFRQPSLESDDEDVDSMGGEDDEDVGD-----BEEGQSE	1357	
Db	1028	-----LPK-----EDIEEHFKKEDSEPLGRIRQKRKYVYDDGLTE	1065	
Qy	1358	BEVEVEQDEDDSOEBEVSULPKKGRQOVRLPVKTRGKLSSPSRSGQQOEPGRYPSRQ	1417	
Db	1066	EQFLEAVEDDNMSLED--AIKKERARERRRLRQNGTKENEI-----ETL	1108	
Qy	1418	QSTPKTTVSSKTRSLRKINSAPTEIKSLR--TASBSTRSHGHPLOADVVELLSPRK	1475	
Db	1109	ENTPEASETSLIENN--SFTAANDVEENTADKETTASRKR-----SRKK	1152	
Qy	1476	RR-----GRKSANNTPENSPNPN-----FRVI--ATKSSFOSSRVNIAKSLS-	1516	

Db 1153 RTISIVTADKENTQESTSQNGGAKVEEVKSSVEIINGSKSKKKPKLTVKIKLNK 1212
Qy 1517 ---LQESKSRCKRKQSPSPVTLGRSSRGQGVHLSAPQLVVL---VRHDDSW 1570
Db 1213 TTVLENNQKRAEPEKSPAKKTAAKTKTKSKSLGIFPTVEKLVEMREQLDEVDSH 1272
Qy 1571 P----FLKLVSKIQVPDYDIIIKKPIALMIIRKVNKCEYKLASEFIDDIELMFNSNCFY 1626
Db 1273 PRTSIFEXKLPSKRDYDPYFKVPEKPAWDIILKCNKNGTKYKLTLEEVQALQTMFENARFY 1332
Qy 1627 NPNRTSEAKGTRLOAFPHIQAKL 1651
Db 1333 NEEG-----SWYVVDADKL 1346

RESULT 41
B43402
myosin heavy chain-B, neuronal - chicken
N;Contains: myosin ATPase (EC 3.6.4.1)
C;Species: Gallus gallus (chicken)
C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 09-Jul-2004
C;Accession: B43402; A43402
R;Takahashi, M.; Kawamoto, S.; Adelstein, R.S.
J. Biol. Chem. 267, 17864-17871, 1992
A;Title: Evidence for inserted sequences in the head region of nonmuscle myosin specific
yosin.
A;Reference number: A43402; MUID:92388144; PMID:1355479
A;Accession: B43402
A;Molecule type: mRNA
A;Residues: 1-2007 <TA>
A;Cross-references: UNIPROT:Q02015; GB:M93676; NID:G212448; PIDN:AAA48988.1; PID:G212452
A;Note: the sequence of residues 212-221 and 632-652 and the corresponding nucleotide se
A;Accession: A43402
A;Molecule type: mRNA
A;Residues: 1-211;722-631;653-2007 <TA2>
A;Cross-references: GB:M93676; NID:G212448; PIDN:AAA48988.1; PID:G212449
A;Note: sequence extracted from NCBI backbone (NCBIN:112864)
C;Comment: Alternatively spliced segments 1 and 2 are found exclusively in nonmuscle myo
C;Superfamily: myosin heavy chain; myosin motor domain homology
C;Keywords: actin binding; alternative splicing; ATP; coiled coil; hydrolase; methylated
F;1-2007/Product: myosin heavy chain-B, neuronal #status predicted <MYN>
F;1-211,722-631,653-2007/Product: myosin heavy chain-B, nonmuscle #status predicted <MYN>
F;88-802/Domain: myosin motor domain homology <MMOT>
F;178-185/Region: nucleotide-binding motif A (P-loop)
F;212-221/Region: alternatively spliced segment 1 #status experimental
F;559-593/Region: actin binding #status predicted
F;632-652/Region: alternatively spliced segment 2 #status experimental
F;692-714/Region: actin binding #status predicted
F;875-2007/Domain: coiled coil #status predicted <COI>
F;875-1315/Region: S2
F;1316-2007/Region: light meromyosin
F;129/Modified site: N6,N6-trimethyllysine (Lys) #status predicted
F;184/Binding site: ATP (Lys) #status predicted
F;732,742/Active site: Cys #status predicted
F;1954/Binding site: phosphate (Thr) (covalent) #status predicted
F;1987/Binding site: phosphate (Ser) (covalent) #status predicted

Query Match 3.11%; Score 271; DB 1; Length 2007;
Best Local Similarity 20.4%; Pred. No. 0.0004;
Matches 297; Conservative 210; Mismatches 484; Indels 464; Gaps 70;

Qy 326 FQNGK-----KKDAIDPLPKYKVPQTK-----KELHESAIYKATQISRKHLS 370
Db 771 FMDGQACERMIRALEDPNL--YRIGQSKIFFRAGVLAHLBERDLUKITDIIFQAVC 828
Qy 371 RDKL--KLFLKHQHCPEQGVIKKASSLSTY-KIAEQDFSFFPDPTTFSPANRRR 427
Db 829 RGYLARKAFKK--QQQLSALKILOCNCAYLKLHWHQWRF-----TKV 872
Qy 428 RPPKRIHISQEDNVANKOTLASYSRKATK-----ERDKLLKQ----- 464
Db 873 KPLQVTRQBELQAKDELMDVKVKOTKVEAELEBMRKHOOLEKNTLAELOAETE 932


```
QY 740 --GKLLTVS-----TRDFIEDVVDILROAKQEFRELKABQHRKE 777
Db 1228 KRGPILLTALSAAVASALLKLHQLDWKTQOTRDVLRDQVQKLEERLTDTAEKSOVH-TE 1286
QY 778 REEAARIRKREK-----LKEQEQWKKEQKLEKEDORNSTADISIGEE 824
Db 1287 LODLQRLQSQEQEKKWEGKONSLESEMELHETMASLQGRRLRAELQRMQAQ---GER 1343
QY 825 E-----REDFDTISBSKTEQK-----LQDMFT-----EDEDPPGSHKGR 862
Db 1344 ELLOAAKENLTAQVEHLQAQAAVEARAQAAGILEEDLTARSALKLNBEVESERERAAQ 1403
QY 863 ---RGK-----RGQGFKEFTQEQINCVCVTRRELL 888
Db 1404 ALQEQGLKVAQCKALQENLALLTQTLAERBEEVETLRQ--IQLEKQREMKALELL 1461
QY 889 TADEEALQKHORKEKELLEKIOSAIACTNIPPLGRDBMYRYWIFPSICLPFEEDYS 948
Db 1462 SLD-----LKNQNEVDLQ-QBOIQELEKCRS-----VLEHLP-MAVOEREQ 1501
QY 949 GLTEDMLPRPSSFQNNVQSQDPQVSTKTGTPLMSESTSNIDQGRDHSVQLPKPVHKPN 1008
Db 1502 KLT-----VQREIQREPEKDRQTVNLEHQLLEL----- 1531
QY 1009 RWCYSSCEQLDQLEALNSRGHRESALKETLLQEKSRICAQARFSEKTFHSDKPPQD 1068
Db 1532 -----EKXDOMIE--SQRGVQDLKKQLVTLE---C-LALEENHHKME----- 1570
QY 1069 SKPTYSRGRSSNAYDPQSMCAKEKLE-----LALRDLFDIEDRIYQGTILGAIKVTR 1121
Db 1571 -----COQKLIKELQEGRETQRTVALTHLTDLEERSQELQAQSSQIHD- 1613
QY 1122 HIRSALESRGYELLSEENKNGI1KTVNEQVVEEMIDEQTKVIKDRLLG1K1TPTSTV 1181
Db 1614 -----LESHSTVLARELQERDQEVKSQREIEELQROKE----- 1647
QY 1182 STNASTPOSVSVVHYLAMALFQIBQGIERRFLKAPLDASDGRSYKTVLD--RWRESLL 1239
Db 1648 -----HLTDLERRDQELMLQ-----KERIQVLEDRQTRQTKI 1680
QY 1240 SSASISQVFLHLSLTDLSRVISKSTLNARCKIKKGDAAENVLDCDGRG---HHYTCV 1296
Db 1681 LEEDLEQIKLSLRERGLTTQRLQMOERAE--EGKGPSK-----AQRGSLHMKLIL 1731
QY 1297 RPKLTVPEGDWFCPCRPK----- 1316
Db 1732 RDKKEV-----ECQEHHELQELKDQLEQQLGLHKKVGTSLLSQREQEIUVL 1783
QY 1317 ---QRCRLSPRQPSLESD-EDVEDSMGGEDDEVDG-DBEEQOSEBEEYEVEQDEDD 1370
Db 1784 QQQLQEAAREQELKEQSLSQDLDAQRALAQRDQELEALQEQOQAQOEERVKEKADAL 1843
QY 1371 Q---EEEVSLPKR-----GRQVRLPVKTR-----GKLSSSFSGRQGOQEPG- 1410
Db 1844 QGALQEAHMTLKERHGLQDHKEQARRLEELAEVAGREVRVQALEEVGLDLRAESREQEKAL 1903
QY 1411 -----RYSRSQOSTPKTVSSKTGRSLRKINSAPPTETKSIRIASRSTRSH--GPLQA 1463
Db 1904 LALQOQCAEQAEQEHVEFTRALQDSWLQAOAVLKERDQELEALRAESQSSRHOEEAARARA 1963
QY 1464 DVVFELLSPRKRRGRKSNANTPENSFPNFRVIATKSSQSRSVNIATSKLSQESRSK 1523
Db 1964 EALQALGKAHALQOKEQH-----LLEQALSRSLQ-ASTATLQASLDA 2007
QY 1524 RRCRKQSPSPSVTLGRSSRGQGVHLSAFEQLVY---ELVRHDDSWPFL---KLVS 1577
Db 2008 COAHSRQLEALRIQEGEIQDQLRYQEDVQLOQALAQRBELRHQOREQLLEKSLAQ 2067
QY 1578 KIQVPDYDIIKKPTALNIIREK-----VNKCEYKLASEFIDDIELMFSNCFEYNP 1628
Db 2068 RVQ-----ENMIQEQKNTGLEREEREEIRGLHQSREVLQTLAQKEQEILELR-----ETQQ 2118
```

```
QY 1629 RNTSEA 1634
Db 2119 RNNLEA 2124
```

RESULT 43

A64224

hypothetical protein MG218 - Mycoplasma genitalium

C:Species: Mycoplasma genitalium

C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004

C:Accession: A64224

R:Fraser, C.M.; Gocayne, J.D.; White, O.; Adams, M.D.; Clayton, R.A.; Fleischmann, R.D.;

M.; Fuhmann, J.; Nguyen, D.; Utterback, T.R.; Saudek, D.M.; Phillips, C.A.; Merrick, J.;

C.A.; Venter, J.C.

Science 270, 397-403, 1995

A:Title: The minimal gene complement of Mycoplasma genitalium.

A:Reference number: A64200; MUID:96026346; PMID:7569993

A:Accession: A64224

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-1805 <TIGR>

A:Cross-references: UNIPROT:P47460; GB:U39699; GB:L43967; MID:g1045903; PID:g1045905; TI

A:Experimental source: strain G-37

C:Genetics:

A:Genetic code: SGC3

C:Superfamily: Mycoplasma genitalium hypothetical protein MG218

Query Match 3.1%; Score 266.5; DB 1; Length 1805;

Best Local Similarity 18.9%; Pred. No. 0.00055;

Matches 269; Conservative 242; Mismatches 506; Indels 403; Gaps 59;

QY 330 KKDAIDPLLPKYVQPTKELHESAIYKATQISRKHLSRDKLKLKLF---LKQCEPQE 386

Db 473 KKVDEAIFQLKEKVAQERKELEELYLVK---KQKQDKENELLFFEKQLKQADFE 527

QY 387 GVIKIKASSLTYKIA-EQDFSYPDPDPPTTFIFSPANRRGRPPKRIHISOEDNVAKQ 445

Db 528 NELSAKQOELPEAKHALERSPI-----KLEDEKDLNTRAQ 563

QY 446 TLASVRSKATVERDK---LLKQEMKSLAFKAKLKREK-----ADALEAKKK 490

Db 564 QIANEFSLTKDSKSDAFELMLQNEYNLOEQEKQLQERTYFERNAAVLSNRLQKRE 623

QY 491 EKEDKKEKREELKIVBEERLKKKEERLKVREKEREKREKREKREKREKREKREKREKRE 550

Db 624 ELLQOKETLDQLTKSFEQERLINQREHKELVASVEKQKELGKQLQDFSQTSLNASKNLA 683

QY 551 DMECDDLKELPEPTPVKTRLPPEIFGDALMVLEFLNAPGELFDLQDEFPDGVTVLESEA 610

Db 684 ERE-----MAIKPKEK-----EIEATBKQ 702

QY 611 LVGNDSGPLCELAFFFLTAIFAEEEEEVAKELTDADTKGCSLAKSLDLSCTLSEI 670

Db 703 LL-NDVNN--AEVIOADLAQLNQSLSQNSQLQNAKQRIADFNHDSLKKLMEYELSLQK- 758

QY 671 LRLHLASGADVTSANAK-VRYQKRGGFDAITD---ACMELR-----LSNPSSLVKKL 718

Db 759 -RLQEL---QTLQANQKHSYQNAQYFEGELDKLNREKQAFNLNRKKQTHEVDAIKQRL 813

QY 719 SST-SVVDLTPGEKMKILH-----ALCGKLLTLVSTRDIEDYVDVILROAK 763

Db 814 SDKHQALNMQAELDRKTHLNNAPLNHDADQKSLQDQLATVKETQKLDLERSALLEKQ 873

QY 764 QEPRELKA--EQRKEREAAAARIRKKEEKLKEQEKMKKEQKELK---EDEQRNSTAD 818

Db 874 REFAPNAVGRHWSNKTSLQKJI-----YELTKQESQTKQKTELKIAFSDLQK-----D 925

QY 819 ISIGEEEREDPDTGIESKDTQEKELDDQMFTEDEDDPGSHKRGRRGKRGKQNGFKFEFTQ 878

Db 926 YQVELQKQOEFQRIEAK---QRELDK---LAEKNQVLELDNRFPQALQKQ-KQDTVQA 978

QY 879 QINCVTRELLTADBEELAKQEHQ---RKEKLELEKIQS-----AIACNTIPL 923

Db 923

Db 1028 KRHEKLTLDRLAEKEKDHKDAEINORFKQFENEYADFQAKKRELQELNQRNLE 1087
QY 1081 AYDPSOMCAEQLEL-----RLRDFLLDIEDRIYQGTGLGAKVT 1119
Db 1088 QSNASLLKRNQTLDFALLRKVQHTQTNRVOLNTQIKEFLL--EKNQF-----KAS 1139
QY 1120 DRHIWRSALSGRYELLSENKENGIIKTVDNEVEMEIDEQTKVVK---DRILGKITE 1176
Db 1140 DEAAQLQAL-----LIRKLSRFSKQLQLOREALAIQKLEFDK---RDE 1179
QY 1177 TPSTVSTNASTPQSVSVVHYLAMALFOIEQ-----GIERRFLK 1215
Db 1180 QOKSEINNAK-----LQLQFKLEKQFNDEAKQKQLEIFKQOQORLDVKEKRLK 1228
QY 1216 APL-DASDGRGYKTVLDRWRESLLSSASLSQVFLHLSTLDRSVIWSKSIINARCKI 1274
Db 1229 QKLVQLKNSKSYLVYKNR---ADLSQQOLQHKYANLLEKELQAKRAL----- 1276
QY 1275 KGDAENMWLDCDGRGHHTYCVRPKLKTVPEGDWFCPCRPKQRCRRLSFRQPSLESDE 1334
Db 1277 -----DKKGR--AIYGMQWQ-----FVSELQRE-----KKQLLSAQK 1306
QY 1335 DVEDS---MGGRDDEVDGDEEGQSEEEYEVEQDEDDSOEEVEYS-----LPRKG 1382
Db 1307 QVDDKSRLLQONRHLNLSSTKTKROSLEHDINKFDORKEAVSSILNSHKLKQKEG 1366
QY 1383 RPQ---VRLPVKTRGKLSSFSRQOQEPGRYPSRQOQSTPKTTVSSKTRSLKINS 1439
Db 1367 ELQILQKLSLK-KTOIEQEFKLYQRE-----KLDQRRTL-SKLHRELKAQNEA 1416
QY 1440 PPTETKSLRIASRSTRHSHGFLQADVVELLSPRKRRGRKSANTP-ENSPN--FPNFR 1496
Db 1417 -----TAHKREY-----LEIENYKKEQLQRLTKSEFDDNNKRLUFYER 1457
QY 1497 VIATKSSBOSRSWNIAKSLQESKRRCKRQSPSPVTLGRSSRQSGVHLSAP 1556
Db 1458 KIRNIEKKEAHI-----KTVLETKKKHLVETAVK---LHLQKQSIISKG--QELKEI 1508
QY 1557 EQLVVELVRH 1566
Db 1509 KERSVRDISH 1518

RESULT 45
JC5420
smooth muscle myosin heavy chain 1 - mouse
C:Species: Mus musculus (house mouse)
C:Date: 10-Jun-1997 #sequence_revision 18-Jul-1997 #text_change 09-Jul-2004
C:Accession: JC5420
R:Haegawa, K.; Arakawa, E.; Oda, S.; Matsuda, Y.
Biochem. Biophys. Res. Commun. 232, 313-316, 1997
A:Title: Molecular cloning and expression of murine smooth muscle myosin heavy chains.
A:Reference number: JC5420; MUID:9742182; PMID:9125171
A:Accession: JC5420
A:Molecule type: mRNA
A:Residues: 1-1972 <HAS>
A:Cross-references: UNIPROT:008638; DDBJ:D85923; NID:g1945077; PIDN:BAA19690.1; PID:g194
A:Experimental source: smooth muscle
A:Comment: This protein plays a role in smooth muscle cell contraction.
C:Superfamily: myosin heavy chain; myosin motor domain homology
C:Keywords: nucleotide binding; P-loop
F:188-771/Domain: myosin motor domain homology <MMOT>
F:178-185/Region: nucleotide-binding motif A (P-loop)

Query Match 3.1%; Score 266.5; DB 2; Length 1972;
Best Local Similarity 19.6%; Pred. No. 0.00061;
Matches 230; Conservative 207; Mismatches 415; Indels 319; Gaps 49;
QY 212 LTYQALESEKARQNSQFPEPIIPVLYLTSLTHRSRLHEICDDIEA-----YVKD 264
Db 953 LDLEQLSEEEAARQKQ-----LEKVTAEAKIKKLEDDILVMDQNSKLSKE 1000
QY 265 RYFVEETVEIRNNGARLQ---CTILEVLPPSHQNGFANGHVNSVDGETIISDSDSETQ 322

Db 1001 RKLLEERYSDLTNTLAESEEKAKNLTCLKSKH-----ESMI-----SELE 1040
QY 323 SCSEFQNGKKDAIDPLLFKYVOPPKKELHSAIVKATQISRRKHLFSRDKLKLFL-KQH 381
Db 1041 VRLKKEESRQLEKL--KRLKLGASDFHQIADLQAOIA-----ELKQOLAKKE 1089
QY 382 CEPOGVITIKASSLSTYKIAEQDPSFYFPDPPTFFISPNRRRGRPPKRI-----HIS 436
Db 1090 EELQAALARLDE-----EIAQKNAL-----KKIRELEGHIS 1121
QY 437 --QEDNVANKOTLASYSRKATKERRDKLLKQSEMKSIAFEKAKLKREKADALEAKKEXED 494
Db 1122 DLQEDLDSER-----AARNKAQKQ-----RDLGEBELAKLTEDETLDDSTATQOEL 1168
QY 495 KEKKREE---LKKIVVEERLKKKEERLKVVEREKEREKREKREKRYVEYLKQWSPKRED 551
Db 1169 RAKREQEVTVLKKALDESTRSHEAQVQEMRQKHQAVEELTEQ-----LEQFRAKAN 1221
QY 552 MECDDLKELPPTPVKTRLPPEIFGDALMWLEFLNAFGE--LFLDQDEFFPDGV----- 602
Db 1222 L--DKSKQTLKENADLAGELRVLGQAQOEVEHKKKLEVLQDLQSKSCDGERARAEIS 1279
QY 603 ----TLEVLLEALVG--NDSEGPLCELLFFFLTAIFAIAEBEEBEVAKEQLTDATKGS 656
Db 1280 DKVHKLQNEVESVTGMLNEAEGKAIKLAKD--VASLGSQLODTQELLOEETRQKLNVS TK 1337
QY 657 LKSLDLDSCITSEIL-----RLHILASGADVTSANAKYRKRGGFDDATDDACMELR 708
Db 1338 LRQLEDERNSLODQDDEMEAKQNLERHVSNTLNTQLSDSKKLQ--DFASTIEVWE-- 1392
QY 709 LSNPSLVKLLSSTS-----VYDLTPGEKMKILHALCGKLLTLVSTDRFIEDYVDILR 760
Db 1393 -GKKRLQKEMEGLSQOYEEKAAAYDKLEKTKNRLQOELDLVVDLDNQRLQVSN---LE 1447
QY 761 QAKQEPRELKAEQ-----HRKEREBAARIRKK-----BEKLEKEQ 798
Db 1448 KKQKFPQLLAEKKNISSKYADERDRAEAAREKETKALSALARALEALEAKEELERTNK 1507
QY 799 KMKKEQKEL---KEDEQRNSTADISIGEEBEDPDTSESQTEOKELDDQM-FTED--- 851
Db 1508 MKKAMEDLVSSKDDVGKN-----VHELEKSKRALETQWEEKMTQLEESDDVQATEDAKL 1563
QY 852 -----EDDPGSHKRRGRKGQNGFKFETRQEQINCVTRELLTADEEAL-----KOE 899
Db 1564 RLEVNMQALKQGFERDLQARDEQNEEKERQQLRQLHEVETELEDERKQALAAAKKLE 1623
QY 900 HORKEKELLEKIQAIACTNIFPLGRDRMYRYYWIFPSIPGLFIE-EDYSGLTEDMLLPR 958
Db 1624 GDLKDEL--QADSAIK-----GREEAIKQ-----LRKLAQMKDFORELDDARASR 1668
QY 959 PSSFQNNVQSDPQVSTKTGTGEPMLSESTSNIDQGRHVSQVLPKPVHKPNRWCFVSSCEQ 1018
Db 1669 DEI FATS KENKKA KSL EADLMQLQEDLAAERARKQADLEK-----EL 1713
QY 1019 LDQLIEALNSGRHRESALKMETLLQKSRICQAQLARFSEKPFH-----FSDKFPQPSKP 1071
Db 1714 AEELASSLSGR-----NTLODEKRLLEARIAQLBEELEEEQGNMEASDRVR---KA 1762
QY 1072 TYSGRSSNAYDPSQMCABK-----QLELR--LRDFLLDIED----- 1106
Db 1763 TLQAEQLSNELATERSTAQKNESARQQLERQNKELRSKLQVEGAVKAKLSTVAALBAK 1822
QY 1107 -----RIVQGTLGAIKYVTRDRHWRSALESGRYELLSENKE-----NGIITKV 1149
Db 1823 IAQLHEQVEQAREKQKQATSKLQKQKLEKVLQVEDERKMAQYQKAEKGNKTVKQL 1882
QY 1150 NEDVEEMEIDQTKVIVKDRLLGKTKETPTSTVSTNASTPQSVSVVHYLAMALFOIEQGI 1209
Db 1883 KRQLEAE--EESQCINANR-RKLORELDDEATESNEANGREVNALKS-----KLREGN 1932
QY 1210 ERRFL-----KAPLDASDS 1223

RESULT 47
T22976
hypothetical protein F59A2.6 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T22976; T23157
R:Lightning, J.
submitted to the EMBL Data Library, June 1994
A:Reference number: Z19645
A:Accession: T22976
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1133 <W1>
A:Cross-references: UNIPROT:Q21022; EMBL:Z34801; PIDN:CAA84332.1; GSPDB:GN00021; CESP:F59A2
A:Experimental source: clone F59A2
R:Burton, J.
submitted to the EMBL Data Library, October 1995
A:Reference number: Z19700
A:Accession: T23157
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1133 <W12>
A:Cross-references: EMBL:Z66514; PIDN:CAA91344.1; GSPDB:GN00021; CESP:F59A2.6
A:Experimental source: clone K01A11
C:Genetics:
A:Gene: CESP:F59A2.6
A:Map position: 3
A:Introns: 13/2; 43/3; 107/3; 413/3; 492/3; 567/3; 635/3; 710/3; 738/3; 795/2; 1008/3; 1

Query Match 3.1%; Score 265.5; DB 2; Length 1133;
Best Local Similarity 19.6%; Pred. No. 0.00035;
Matches 251; Conservative 183; Mismatches 425; Indels 423; Gaps 55;

QY 442 ANKOTLASYSKAT-----KERDKLK-----QEEMKSLAFKAKLKRKADALEAKK 489
DB 16 AKKALAEKCELTUKFOADKEKKNEMVQSLRQEM-----LSKCDALQAEV 64
QY 490 KEK-----EDKEKKEELKKIVEERLKKKEERLKKVREKER-----EK 530
DB 65 NEAKALREEIQAKYDDVTQKABRIQGLEESKKVLESEKQAFENEKEQEREQELAKANEK 124
QY 531 LREEKRVVYLKQSKPREDMECDLDELPEPTVKTRLPPEIPGDALMVLFLNARGE 590
DB 125 LNSEQNILDVTKLEQSEE-----EVLAAARGA 152
QY 591 LFDLQDEFDPDGVTLVLEALVGNDSGCLLFFLTAIQAIAEEVEEVAKEQLTDA 650
DB 153 IQEL-----TEKJEE-----SEKETSTAKTELEAV 177
QY 651 DTGCSLKLSDLSCTLSEILRLHLASGADVTSANAKRYQKRGGFDDADDACMLRLS 710
DB 178 S-----KKLDSSETSLKEP-----SDMIEA---MKQLINCEKQKDEAVELLKQK 219
QY 711 NPSLVKGLSSTSVYDLTPCEKMKILHALCGKLLTLIVSTRDFIEDYVDI-----LROAK 763
DB 220 LEEVEKNMSDVEVQ-----KQLLESTTSEMKQHAEEAATVKKQLEBAQ 263
QY 764 QEFRELK--AEQHRK-----EHEEAAA--RIRKKEEKLKQEQKMKQEP-----805
DB 264 SSIEHLKKAENERNLKTALDESDESSAISSETTKQMAAKKELEAASEKESLREQMDRLQ 323
QY 806 -----KLKEDQ-----RNSTADISIGEEREDFD 830
DB 324 KVHNAQBEDIQKLTWELEWAKIAKSTEDKLAEREQAGLENAKEDLKVVEEKH-----380
QY 831 TSIESKD-----TEQKELDQDMFTEDDEDPGSHKRGH---RGRKGNGFKFEX 874
DB 381 TGIQRAQALDDAEKVEKVLQELERAQSALESSQELASSQKADKIQLELEKELQNAQKS 440
QY 875 -----TROEQNCVTVRELLTADDEE--EALQKHQRKEKELLEKIQSAIACTNIFPLGRDM 928
DB 441 SEELATANEMVRSLTATLNSNSSETIILKQLETLDELQARQQOTEKALT-----490

QY 929 YRRYWI PPSIPGLPIEDYSGLTEDMLLPSPSPQ--NNVQSDQDPQVSTKTGEPL-----981
DB 491 -----EEINVLTTSLAEKEQQTAAIQNLQTQIYQMEVEKEKVELVKVQ 534
QY 982 ---MSESTSNIDQPRDHSVQLPKPVHKPNCWCFYSCEQLDQLIEALNSGRHRESALKE 1038
DB 535 LQQAQSSSSAAEEALRAEIQLEAKL-----KAVEQAK--AAELNS-----573
QY 1039 TLLQEKSRICQAQLARFSEKHFSDKPPQSDKPTYSRGRSSNAYDPQSQMCAEKOLELRLR 1098
DB 574 -LLAEKEHLQALHQLGVKE---BEKLEWVKVQLQQAQSSSSVEQALRAIEKLEAKLQ 629
QY 1099 DFLDLIEDRIYQGTGALKVTDRHWRALSBSGRYELLSEENKENGHIKTYNEDVEEME 1158
DB 630 ---EIEEE-----KKNALNAS---LAEKEQQTAAIQLEQAQLHQLLEV 665
QY 1159 DEQTKVIVKDBLLGIKTETPTSTVSTNASTPOSVSVVHYLAMALFOIQEGTERRFLKAPL 1218
DB 666 EKEEKLME-----VKVQLQQAQSSSSVEQALRAEIEKLEAKLQETB-----KAKM 711
QY 1219 DASDSG---RSYKTVLDRWR--ESLLSSASLSQVPLHLSTLDRSVINWSKSLNARCKICR 1273
DB 712 QNSSKREQKVRLESLNLEKVRVEFIAKEKILSDRLSELSTISTELVVOKATVE---KTKM 768
QY 1274 KKGDAENMVLCDGCDRGHTTYCVRPKLKTVPEDWFCPECRPKQRCRLSPFRQPSLES 1333
DB 769 DFGELE-----TREKRATAD-----RENEKMBEIRLURTFPAKEL- 802
QY 1334 EDVDSMGG--EDDEVGDDEEGOSE-----BEEYVEQDEDDSQEVEEVSPLPKR 1381
DB 803 ---ETMGSALVEKETAYNELKASAEKKIAKLTQSQFEKLLKKAQNSQDEASESRFKTLEA 858
QY 1382 GRPOVRL-----PVKTRGKLSSSPSSRGQO---QSGRYPSPSSQOSTPTKTVSSK---TG 1430
DB 859 SAEQAKLESEQKRALBELLSKSESEIEELKIKISAEKRSKSHVEKEMLEGEAKEULTD 918
QY 1431 R-----SLRKINSAPPTETKSLRIASRSTRSHSGPLQADV---FVELLSP-----RR 1474
DB 919 RIEGLEAEVKKLTAA--NETKAVK--ADTDARKVVRELQKEVKQLYNELNDKNOQFDMVQE 975
QY 1475 KRRGRKANNTPNSPNFPNFRVIATKSSSEGRS-----VNIAKLV---SLOESESQR 1524
DB 976 ELTKLTKSKETAENG-----QLQVQKQMDDEDRRSEFSEFKIEIASLKQKLDASLTADDLR 1031
QY 1525 RCRKQSPSPVTLGRSSGR 1546
DB 1032 MQVSRN--EKTFRSGNASTGR 1051

RESULT 48

F96712

hypothetical protein F14K14.10 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004

C:Accession: F96712

R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,

ansen, N.F.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Dewar, K.;

Chin, C.W.; Chung, B.; Huizar, L.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719; PMID:11130712

A:Accession: F96712

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1085 <STO>

A:Cross-references: UNIPROT:Q9CA42; GB:A8005173; NID:96524182; PIDN:AAFI5067.1; GSPDB:GN

C:Genetics:

A;Gene: F14K14.10
A;Map position: 1

Query Match	3.1%	Score 264.5	DB 2	Length 1085
Best Local Similarity	20.7%	Pred. No. 0.00037		
Matches 240	Conservative 163	Mismatches 395	Indels 361	Gaps 55
Qy	172	EEVYCKVTWEIFRHYDDFFB--RTILCNVLWSCAVTGRPLGTIOEALSEKKAQNQL	228	
Db	87	EKKQWTSNNELQAYDEAMEMLKREKTSNA-----ITLNEADKRENLKAL	134	
Qy	229	---QSFPPELIPVLY-----LTLSTHRSRLHE-----ICDDIFAYVKDRIYFVBETV	272	
Db	135	IDEXQFVAELENLDKYQREHSVVKSTSEAKLEBANALVIGMKEKALEVDREIRAIAEKF	194	
Qy	273	EVIRNNGARLOCTTILEVLPSPHQNGFANGVNSVD-----GETIIISDDSDSETOSCSFQN	328	
Db	195	SVNMRKSSELERKUKEV--ETREKVHQREHLSLVTEREAHEAVFYKQRED-----LQE	245	
Qy	329	GKKKDAIDPLLKYKVQPTKKELH--ESATVKATQISRRKHLFSRDKLKLFL--KQHCEP	384	
Db	246	WEKLTLE-----EDRLSEVRKSIHNRERVMENERTIEKKEKILIENLQOKISVAKSELTE	301	
Qy	385	QEGVIKIKASLSYTKIAEDQFSYFFPDPPPTFIFSPANRRRPPPKRIHIISOEDNVANK	444	
Db	302	KEESIKIKLNDIS--LKXD-----FEAMKAVDIKELHEFEENLIERE	345	
Qy	445	QTLASYRSKATKRDKLLKQEE-----MKSLAFEKAKLREKADALEAKKE-----	491	
Db	346	Q-----METGKLLDDQKAVLDSRRRREFEMELQMRSLDEELGKKAETLOLQVE	395	
Qy	492	--KEDKEKKRE--ELKIVBERLKKKEE--KERLKVBEREKEREKLEEKRYVEYLKQW	545	
Db	396	ISHKEEKAKREAALEK--KEEGVKKKEKOLDARKLTKVEKBAKALKAEEKKLMHEN----	449	
Qy	546	SKPREDMEC-----DDLKEL--PEPTPVKTRLPEIFGDALMV-----LEFLNAGELFDL	594	
Db	450	ERLLEDKECLRKLADEIEIGTETTKQESRIREE--HESLRITKEERVEFLR-----L	500	
Qy	595	QDEFPDGVTVLEALVGNDSQPLCELLFFFLTAIFAQIAEBEEVAKEQLTADTKG	654	
Db	501	QSELKQIQIDKVKQBEELLKERE-----ELUKDKERFEK-----	535	
Qy	655	CSLKSLLDSCITLSEILRLHLTASGADVTSANAKRYQKRGGFATDDACMELSLNPSL	714	
Db	536	--WEALDKK-----RANITRQNEVAENEKLR-----NLQISEKHR	570	
Qy	715	VKLSSTSVYDL--TPGEKMKILHALCGKLLTVSTRDFTIEDYVDILRQAKQBFRELKA	771	
Db	571	LKREEMTSRDLNKRELDGVKKQ-----KESFEADMEDLEMQRKNDMEFQR	616	
Qy	772	EOHRKERE--EAAARIKREKEKL-----KQEQKQKKEK-----QEK	806	
Db	617	QEEAGERDFNERARTYEKRSQEEELDNINITYTKLAQREEMEQYBKALEREREQISVRKK	676	
Qy	807	LKDEQRNSTADI-----SIGEE-----EREDFDTSIE-----SKDTEQKELD	844	
Db	677	LLKEQEAEMKHDI TELDLVLSLKEKKEKFEFCERERFLVFEUKLSKSSCGEITENFVL--	735	
Qy	845	QDMFTEDEDDPGSKHRRGRKRGONGKFETFRQIQINCVTTRELLTADDEEALK---QEHQ	901	
Db	736	SDLRLPDVED-----GDKRFQKQK-----LKAEEALNISPAAEN	769	
Qy	902	RKEKELLEKQSAIACACTNIPFLGRDRMYRRYWI PFSIPGLFIEDYDGLTDEMLLPRSS	961	
Db	770	SKRTSLGKIASKL--LSISPIGK-----TDKVTDLGITVVLPESSQ	809	
Qy	962	FQNNVQSQDPQVSTKTGEPLMSEST---SNIDQGPDRHSVOLPKPVHKPNRWCFYSCEQ	1018	
Db	810	PDDSLD-----RVSGEDHEPSATEQSF TDSRIQEGP--EGSLQSEMKSKDPRG-----	856	
Qy	1019	LDQLIEALNRSRGH-----RESALKETLLQEK-----SRICAQLARFSEKPHFS	1062	

D_b 857 -----RGRGRRGKSVRSQATVAVSRDSKPSDGTETPRKRQRBEQTSTRITESEQAAG 908
Q_y 1063 DKPQ-PDSPKTYSGRSSNAYDP-SQMCAEQLELRDLRLDFLLDIEDRIYOGTLCAI-----1116
 : ::::::|::|::|::|::|::|::|::|::|::|::|::|::|:
D_b 909 DSDGEVDSTITGGRRKKRKQIAVPVSQTGPQTRYLQRHRNVGTDEKA-QASKGATEKQE 967
 |::|::|::|::|::|::|::|::|::|::|::|::|:
Q_y 1117 -----KVTDRIHWSALBSRGYYELLSBNKENGIIKTWNVEDVEEMIDEOTKVIKDR 1169
 |||||::|::|::|::|::|::|::|::|::|::|::|::|:
D_b 968 RVNDIRKPSPKETRTPPESG-----ENRENGAEVLVTTHTEEI-----1008
Q_y 1170 LLGIKIKTEPTPVTSINASTP 1188
 : ::::::|::|::|::|::|::|::|::|::|::|::|::|::|:
D_b 1009 -VVETEIVFKVANNTGNP 1026
 |::|::|::|::|::|::|::|::|::|::|::|::|:

```

RESULT 49
JC5421
smooth muscle myosin heavy chain 2 - mouse
C:/Species: Mus musculus (house mouse)
C:/Date: 10-Jun-1997 #sequence_revision 18-Jul-1997 #text_change 09-Jul-2004
C:/Accession: JC5421
R:/Haegawa, K.; Arakawa, E.; Oda, S.; Matsuda, Y.
Biochem. Biophys. Res. Commun. 232, 313-316, 1997
A:/Title: Molecular cloning and expression of murine smooth muscle myosin heavy
A:/Reference number: JC5420; MUID:972421B2; PMID:9125171
A:/Accession: JC5421
A:/Molecule type: mRNA
A:/Residues: 1-1938 <HAS>
A:/Cross-references: UNIPROT:O08638; DBJ:D85924; NID:g1945079; PIDN:BAA19691.1;
A:/Experimental source: smooth muscle
C:/Comment: This protein plays a role in smooth muscle cell contraction.
C:/Superfamily: myosin heavy chain; myosin motor domain homology
C;/Keywords: nucleotide binding; P-loop
F:/88-771/Domain: myosin motor domain homology <MMOT>
F:/178-185/Region: nucleotide-binding motif A (P-loop)
```

	Query Match	3.1%; Score 264.5; DB 2; Length 1938;
	Best Local Similarity	19.8%; Pred. No. 0.00073;
	Matches	Conservative 201; Mismatches 406; Indels 295; Gaps 47;

Q_y 212 LTQBALESKKARONLOSPPELTPLVLTSLTHRSLHEICDDIFA-----YVKD 264
 |::|::|::|::|::|::|::|::|::|::|::|::|:
D_b 953 LDLEEQLKEEBAARKLKQ-----LEKVTAEAKIKGLEDDLVMDDQNKSLSKE 1000
 |::|::|::|::|::|::|::|::|::|::|::|::|:
Q_y 265 RYFEETVEVRNNRGARLQ--CTILEVLPSSHONGFANGHWNSVDGETIIISDSDSETQ 322
 |::|::|::|::|::|::|::|::|::|::|::|::|:
D_b 1001 RKLEERVSDLTTNLAEERKANKLYKLXKH-----ESML-----SELE 1040
 |::|::|::|::|::|::|::|::|::|::|::|::|:
Q_y 323 SCSFQNGKKDAIDLPLFYKYVOPTKKELHESAIVKATQISRKHUFSRDKLFL-KQH 381
 |::|::|::|::|::|::|::|::|::|::|::|::|:
D_b 1041 VRLKKEEKSRQEKLQ--XKLGEGADSPHEADIQAQIA-----ELXMQLAkke 1089
 |::|::|::|::|::|::|::|::|::|::|::|::|:
Q_y 382 CEPOEVGKIWKASSLTKTABODFSYFFPDPTTFIFSPANRRRPPIKRI-----HIS 436
 |:|::|::|::|::|::|::|::|::|::|::|::|:
D_b 1090 EEIQALALARLD-----EIQQKNNAI-----KKIRELEGHS 1121
 |::|::|::|::|::|::|::|::|::|::|::|::|:
Q_y 437 --QEDVNANKOTLASVRSKATKERDKLLKQEEKSILAFKAKLKREKADALEAKCKEKED 494
 |::|::|::|::|::|::|::|::|::|::|::|::|:
D_b 1122 DLQEDLDER----AARNKAERKQ-----RDIGEELLEALKTELEDTLDTSTATOEEL 1168
 |::|::|::|::|::|::|::|::|::|::|::|::|:
Q_y 495 KEKKREE---LKKIVEERLUKKEEKERLKVAREKEREKLEREKRYKVEYLKQMSKPRED 551
 |::|::|::|::|::|::|::|::|::|::|::|::|:
D_b 1169 RAKEQEVTVLKALDEETRSHSAQVQMEQKHTQAVEELTEQ-----LEQPKRAKAN 1221
 |::|::|::|::|::|::|::|::|::|::|::|::|:
Q_y 552 MECDDLKELPEPTPVKTRLPETFGDALMVLFELFNAPGE--LFDLQDEFPGDV-----602
 |::|::|::|::|::|::|::|::|::|::|::|::|:
D_b 1222 L-DKSKQTULEKENADLAGELRVLGAOKAQVEHKKKLEVQLQSCKSDGERARAELS 1279
 |::|::|::|::|::|::|::|::|::|::|::|::|:
Q_y 603 ----TLEVLSEALVG--NDSEGPICELLFPFLFTAIPOAIEEEEVAKEOLTADTKGS 656
 |::|::|::|::|::|::|::|::|::|::|::|::|:
D_b 1280 DKVHLQNEVESVTGMINEAGEKAIKLAKO--VASLGSOLQDYELLQEERTROKLNVESTK 1337
 |::|::|::|::|::|::|::|::|::|::|::|::|:
Q_y 657 LKSIDLSDCSCTLSEL-----RLHILASGADVTSANAKRYQKRGGFPAITDDACMELR 708

Db 1338 LRQLEDRNSLOQDDEMEAKONLHSHVSTLNIOQLSDSKKLQ---DFASTIEVWEE-- 1392
Qy 709 LSNPSLVKLSSTS-----VYDLTPGKMKILHALCGKLLTLVSTDFEDYVDILR 760
Db 1393 -GKRLKXMEGLSQYEEKAAYDKLETKRLQOELDDLVDLDNQQLVSN-----LE 1447
Qy 761 QAKQBFREKARQ-----HRKREERAAAIRKK-----EELKKEQEQ 798
Db 1448 KKQKFDQLLABEKNISSKYADERDRAEAARKEKTKALSLARALEALEAKELERTNK 1507
Qy 799 KMKKEQELK---KEDEORNSTADISIGREERDFDTSTESKTEQKELDQDM-FTED--- 851
Db 1508 MLKAEMEDLVSSKDDVGKN-----VHELEKSKGALETQMEEMTKTQLEESDDVQATDAKL 1563
Qy 852 -----EDDPGSHKRRGRKQNGFKETPQEQINCVTRRELLTADAEAL-----KOE 899
Db 1564 RLEVNMQALKGQFERDLQARDQNEKRRQLQRLHVEYTELEDERKQALAAAANKKLE 1623
Qy 900 HORKEKELLEKIQSAIACNIPPLGRDRMYRYWTFPPSIPGLFIE-EDYSGLTEDMLLPR 958
Db 1624 GDLKOLEL---QADSAIK-----GREEAIKQ-----LRKLQAKWKDFORELDDARASR 1668
Qy 959 PSSFNQNVQSDPOVSTKTGPELMSSESTSNIDQPRDHSVQLPKVPKPNRCWCFYSSCEQ 1018
Db 1669 DEIFATSKENEKAKSLEADLMQLEDLAAERARKQADLEKE-----EL 1713
Qy 1019 LDQLTEALNSRSHRESALKETLLOKSRICACALARFSEKHF-----FSDKPPQDSKP 1071
Db 1714 AEELASSLSGR-----NTLQDEKRLLEARIQLEEELESEQNGMEASDRVR---KA 1762
Qy 1072 TVSGRSNAYDPQSMCAEK-----QLELR---LRDFLLDIED----- 1106
Db 1763 TLQAEQLSNELATERSTAKNESARQOLERQNKELRSKLQEVGAVKAKLSTVALEAK 1822
Qy 1107 -----RIYQGTILGAIKVTDRIHWSALESGRYELLSENKE-----NGIITV 1149
Db 1823 IAQLEEQVEQAREQAATSKLQDKKLKEVLLQVEDERKMAEQYKEQAEGKNGTKVKQL 1882
Qy 1150 NEDVEEMIDEQTKVIVKDRLLIGITETPTSTVSTNASTPOSVSV 1194
Db 1883 KQLEEAR---EESQICINNR-RKLQRELDATESNAMGREVNAL 1924
RESULT 50
A33977
myosin heavy chain, nonmuscle - chicken
N:Contains: myosin ATPase (EC 3.6.4.1)
C:Species: Gallus gallus (chicken)
C:Date: 31-Dec-1993 #sequence revision 31-Dec-1993 #text_change 09-Jul-2004
C:Accession: A33977; S06116; A43422
R:Shohet, R.V.; Conti, M.A.; Kawamoto, S.; Preston, Y.A.; Brill, D.A.; Adelstein, R.S.
Proc. Natl. Acad. Sci. U.S.A. 86, 7726-7730, 1989
A>Title: Cloning of the cDNA encoding the myosin heavy chain of a vertebrate cellular my
A:Reference number: A33977; PMID:90046668; PMID:2813355
A:Accession: A33977
A:Molecule type: mRNA
A:Residues: 1-1959 <SHO>
A:Cross-references: UNIPROT:P14105; GB:M26510; NID:g212382; PIDN:AAA48974.1; PID:g212383
R:Katsuragawa, Y.; Yanagisawa, M.; Inoue, A.; Masaki, T.
Eur. J. Biochem. 184, 611-616, 1989
A>Title: Two distinct nonmuscle myosin-heavy-chain mRNAs are differentially expressed in
s.
A:Reference number: S06116; PMID:90032648; PMID:2806244
A:Accession: S06116
A>Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 716-1008 <KAT>
A:Cross-references: GB:X17589
A>Note: this translation is not annotated in GenBank entry GCMHCFMHA, release 114
R:Hodge, T.P.; Cross, R.; Kendrick-Jones, J.
J. Cell Biol. 118, 1085-1095, 1992
A>Title: Role of the COOH-terminal nonhelical tailpiece in the assembly of a vertebrate

A:Reference number: A43422; PMID:92381096; PMID:1512291
A:Accession: A43422
A>Status: nucleic acid sequence not shown; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1900-1959 <HOD>
A:Experimental source: brush border
A>Note: sequence extracted from NCBI backbone (NCBIP:111947)
C:Superfamily: myosin heavy chain; myosin motor domain homology
C:Keywords: actin binding; ATP; coiled coil; hydrolase; methylated amino acid; nucleotid
F:84-764/Domain: myosin motor domain homology <NMOT>
F:174-181/Region: nucleotide-binding motif A (P-loop)
F:552-565/Region: actin binding #status predicted
F:626-640/Region: actin binding #status predicted
F:837-1936/Domain: coiled coil #status predicted <COI>
F:837-1277/Region: S2
F:1278-1959/Region: light meromyosin
F:1937-1959/Domain: carboxyl-terminal <CBT>
F:125/Modified site: N6,N6,N6-trimethyllysine (Lys) #status predicted
F:180/Binding site: ATP (Lys) #status predicted
F:694,704/Active site: Cys #status predicted
Query Match 3.1%; Score 264.5; DB 1; Length 1959;
Best Local Similarity 19.0%; Pred. No. 0.00074;
Matches 262; Conservative 218; Mismatches 417; Indels 481; Gaps 65;
Qy 370 SRDKLKLKQHCPEQGVKIKIKASSLSTYKIAEQDFSYFFPDPTTFIFSPANRRGRP 429
Db 841 SRQEEEMAK-----EELIKVKEKQLA-----EMETP----- 874
Qy 430 PKRHISQEDNVANKQL-----ASYRSKATKRDKL-LKQEMKSLAF-----KAKUK 478
Db 875 -----QAQIMAEKMQLOEQAEALCAEAETIARLTAKQELERICHDLARVBE 927
Qy 479 REKADALEAKK-----EKEDKKEELAKKIVEEERLKKK-----EKER 519
Db 928 EERCHLQAEKKKMQONIQLEEEESARQKLEKVTTEAKKLEEDVIVLEQDN 987
Qy 520 LKVEREK-----BREK---LREKKRY-----VEYLKQMSKPREDME 553
Db 988 LKLAKEKLLDRMSEFTNTLTEEEKSKSLAKLKNKHEAMITDLEERLREKQKQLE 1047
Qy 554 CDDLKELPEPTVTRLPPEIFGDALMVLEFNAPGELFDLOQEPD-----GVTL 605
Db 1048 -----KTRKLE-----GDSSLDHQIAELQAQIAELKIQLS 1079
Qy 606 VLEALVGNDSQGLCELLFFFLTAIFQAIIEEVEEVAKEQLTDADTKGCSLKL-DLDS 664
Db 1080 KKEBELQA-----ALARVEEAAQ-----KNMALKKITRELE- 1110
Qy 665 CTLSEILRLHLASGADVTSANA---KYRYQKRGGFATDDACMELRLSNPSLVKLSST 721
Db 1111 ---SQITELQ-----EDLESERASRNKAKQKR-----DLGELEALKTELEDTLST 1155
Qy 722 SVYDLTPGKMKILHALCGKLLTLVSTDFEDYVDILQAKQEPRELKASQHRKERBEA 781
Db 1156 AAQ-----ELRSKREQEVTLKTKLEDEAKTHEAQIQEMROKH 1194
Qy 782 AARTRKKEEKLKEQEQKKEQKLEK---EDEQRNSTADISI-----GEEEREDPT 831
Db 1195 SQALEELAEQ---LEQTKRVKANLEKAKQALESEAEUSNEVKVLLQGGDAEHKHKVDA 1252
Qy 832 SIES---KDTF-----QKELQDMFTEDDDPGSHKRRGRKRGKQNGFK 873
Db 1253 QLQELQVKFTGERVKTLEAERVNKLQVELDNVTGLLNQSDSKSLA-----KD 1302
Qy 874 FTROEQINCVTRRELL-----TAEDEEALKQEHQKKE----- 905
Db 1303 FSALESQLODQTOELQEBETRLKLSFSFKLQTEDEKNALKEQLEEEAEKRNLEKQISVL 1362
Qy 906 -----ELLEKIQSAIACNIPPLGRDRMYRYWTFPPSIPGLFIEDYSGLTEDMLLPPS 960
Db 1363 QQQAVKARKMDGGLGCLTEAEEAKK-----LQKDESLTQ----- 1399

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 9, 2005, 14:24:09 ; Search time 112.939 Seconds
(without alignments)
4877.138 Million cell updates/sec

Title: US-10-702-148-1
Perfect score: 8631
Sequence: 1 MEDASGRVAPLNNVL.....VTPSNVDQVTPPAKXSR1 1674

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1391452 seqs, 329044822 residues

Total number of hits satisfying chosen parameters: 1391452

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 65 summaries

- Database :
- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
 - 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
 - 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
 - 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
 - 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
 - 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
 - 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
 - 8: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
 - 9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
 - 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
 - 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
 - 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
 - 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
 - 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
 - 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
 - 16: /cgn2_6/ptodata/1/pubpaa/US10D_PUBCOMB.pep.*
 - 17: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
 - 18: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pep.*
 - 19: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
 - 20: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	8631	100.0	1674	9 US-09-839-479-1	Sequence 1, Appli
2	8631	100.0	1674	15 US-10-376-537-1	Sequence 1, Appli
3	8631	100.0	1674	15 US-10-702-148-1	Sequence 1, Appli
4	8588.5	99.5	1673	9 US-09-839-479-69	Sequence 69, Appl
5	8588.5	99.5	1673	15 US-10-376-537-70	Sequence 70, Appl
6	8588.5	99.5	1673	15 US-10-702-148-69	Sequence 69, Appl
7	1215	14.1	279	15 US-10-264-237-2629	Sequence 2629, Ap
8	1087	12.6	211	9 US-09-839-479-62	Sequence 62, Appl
9	1087	12.6	211	15 US-10-376-537-63	Sequence 63, Appl
10	1087	12.6	211	15 US-10-702-148-62	Sequence 62, Appl
11	930.5	10.8	1531	9 US-09-839-479-29	Sequence 29, Appl
12	930.5	10.8	1531	15 US-10-376-537-29	Sequence 29, Appl
13	930.5	10.8	1531	15 US-10-702-148-29	Sequence 29, Appl

14	929.5	10.8	1527	9 US-09-839-479-27	Sequence 27, Appl
15	929.5	10.8	1527	15 US-10-376-537-27	Sequence 27, Appl
16	929.5	10.8	1527	15 US-10-702-148-27	Sequence 27, Appl
17	926	10.7	1525	9 US-09-839-479-68	Sequence 68, Appl
18	926	10.7	1525	15 US-10-376-537-69	Sequence 69, Appl
19	926	10.7	1525	15 US-10-702-148-68	Sequence 68, Appl
20	726.5	8.4	1972	9 US-09-839-479-21	Sequence 21, Appl
21	726.5	8.4	1972	15 US-10-376-537-21	Sequence 21, Appl
22	726.5	8.4	1972	15 US-10-702-148-21	Sequence 21, Appl
23	704	8.2	1969	9 US-09-839-479-71	Sequence 71, Appl
24	704	8.2	1969	15 US-10-376-537-72	Sequence 72, Appl
25	704	8.2	1969	15 US-10-702-148-71	Sequence 71, Appl
26	688.5	8.0	1873	13 US-10-087-192-666	Sequence 666, App
27	682.5	7.9	1876	9 US-09-839-479-70	Sequence 70, Appl
28	682.5	7.9	1876	15 US-10-376-537-71	Sequence 71, Appl
29	682.5	7.9	1876	15 US-10-702-148-70	Sequence 70, Appl
30	680	7.9	1586	13 US-10-087-192-663	Sequence 663, App
31	671.5	7.8	1878	9 US-09-839-479-13	Sequence 13, Appl
32	671.5	7.8	1878	15 US-10-376-537-13	Sequence 13, Appl
33	671.5	7.8	1878	15 US-10-702-148-13	Sequence 13, Appl
34	662	7.7	149	14 US-10-195-730-317	Sequence 317, App
35	662	7.7	149	15 US-10-799-747-317	Sequence 317, App
36	361.5	4.2	1398	15 US-10-094-466-30	Sequence 30, Appl
37	361.5	4.2	1400	9 US-09-764-176-7	Sequence 7, Appl
38	341	4.0	638	16 US-10-437-963-133671	Sequence 133671,
39	318.5	3.7	572	9 US-09-864-761-34546	Sequence 34546, A
40	314	3.6	560	9 US-09-864-761-37938	Sequence 37938, A
41	313	3.6	59	9 US-09-839-479-37	Sequence 37, Appl
42	313	3.6	59	9 US-09-839-479-49	Sequence 49, Appl
43	313	3.6	59	15 US-10-376-537-38	Sequence 38, Appl
44	313	3.6	59	15 US-10-376-537-50	Sequence 50, Appl
45	313	3.6	59	15 US-10-702-148-37	Sequence 37, Appl
46	313	3.6	59	15 US-10-702-148-49	Sequence 49, Appl
47	310.5	3.6	1820	15 US-10-335-977-7982	Sequence 7982, Ap
48	309.5	3.6	1819	15 US-10-335-977-7981	Sequence 7981, Ap
49	308	3.6	1988	16 US-10-437-963-108019	Sequence 108019,
50	305	3.5	2052	16 US-10-437-963-137285	Sequence 137285,
51	303.5	3.5	1130	15 US-10-369-493-6751	Sequence 6751, Ap
52	297.5	3.4	3907	14 US-10-171-311-2	Sequence 2, Appli
53	297.5	3.4	3911	15 US-10-370-685-100	Sequence 100, App
54	297.5	3.4	3911	16 US-10-408-765A-1839	Sequence 1839, Ap
55	297.5	3.4	3925	14 US-10-171-311-6	Sequence 6, Appli
56	295.5	3.4	1875	15 US-10-369-493-22285	Sequence 22285, A
57	295.5	3.4	3899	14 US-10-171-311-4	Sequence 4, Appli
58	295.5	3.4	3917	14 US-10-171-311-8	Sequence 8, Appli
59	292.5	3.4	263	15 US-10-424-599-225426	Sequence 225426,
60	291	3.4	1743	10 US-09-882-227-624	Sequence 624, App
61	288.5	3.3	2468	16 US-10-755-889-615	Sequence 615, App
62	286	3.3	1790	15 US-10-369-493-1586	Sequence 1586, Ap
63	285.5	3.3	1927	15 US-10-282-122A-58729	Sequence 58729, A
64	283.5	3.3	2364	15 US-10-205-331-66	Sequence 66, Appl
65	280.5	3.2	2503	16 US-10-828-985A-11	Sequence 11, Appl

ALIGNMENTS

RESULT 1
US-09-839-479-1
; Sequence 1, Application US/09839479
; Publication No. US20020039779A1
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 08501-042002
; CURRENT APPLICATION NUMBER: US/09/839,479
; PRIOR FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24

```
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1674
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-839-479-1

Query Match      100.0%; Score 8631; DB 9; Length 1674;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1674; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MEDASESSRGVAPLINNVVLPGSPUSLPSVSTGCKSHRVANKVKEARSEKLLPTALPPE 60
DB 1 MEDASESSRGVAPLINNVVLPGSPUSLPSVSTGCKSHRVANKVKEARSEKLLPTALPPE 60

QY 61 PKVDQKLPRSSERRSGGGTGFPPARSRAVAAGEAAAGAAPERGSPGRRVSPRCLCSG 120
DB 61 PKVDQKLPRSSERRSGGGTGFPPARSRAVAAGEAAAGAAPERGSPGRRVSPRCLCSG 120

QY 121 EGGQVAVGVIAKGRGRRGDSRRAPGGREMPLLHRKPFVRQKPPADLRPDDEEVPYCKVT 180
DB 121 EGGQVAVGVIAKGRGRRGDSRRAPGGREMPLLHRKPFVRQKPPADLRPDDEEVPYCKVT 180

QY 181 NEIFRHYDDPFERTILCNLSWCAVTCORPGLTYQEALESEKARONLQSPPEPLIIPVL 240
DB 181 NEIFRHYDDPFERTILCNLSWCAVTCORPGLTYQEALESEKARONLQSPPEPLIIPVL 240

QY 241 YLTSLSHRSRLHEICDDIFAYVKDQRYFVEETVEVIRNNGARLQCTILEVLPPSHONGFAN 300
DB 241 YLTSLSHRSRLHEICDDIFAYVKDQRYFVEETVEVIRNNGARLQCTILEVLPPSHONGFAN 300

QY 301 GHVNSVDGETIIISDDSDSETQSCSFQNGKKKDAIDPLLFYKVQPTKKELHESAIVKAT 360
DB 301 GHVNSVDGETIIISDDSDSETQSCSFQNGKKKDAIDPLLFYKVQPTKKELHESAIVKAT 360

QY 361 QISRKHILFSRDKLFLKQHCPEQGVIKIKASSLSYTKIAEQDFSYFFPDPTFTFS 420
DB 361 QISRKHILFSRDKLFLKQHCPEQGVIKIKASSLSYTKIAEQDFSYFFPDPTFTFS 420

QY 421 PANRRRGPPKRIHISQEDNVANKOTLASYSKATKRDKLLKQEMKSLAFKAKLARE 480
DB 421 PANRRRGPPKRIHISQEDNVANKOTLASYSKATKRDKLLKQEMKSLAFKAKLARE 480

QY 481 KADALEAKKXKEDKKEKREELKKITVEBERLKKKEERLKYVEREKEREKLEERKKYVE 540
DB 481 KADALEAKKXKEDKKEKREELKKITVEBERLKKKEERLKYVEREKEREKLEERKKYVE 540

QY 541 YLKQWSKPREDMECDDDLKELPEPTPVKTRLPPEIFGDALMWLEFLNATGELFDLQDEFPD 600
DB 541 YLKQWSKPREDMECDDDLKELPEPTPVKTRLPPEIFGDALMWLEFLNATGELFDLQDEFPD 600

QY 601 GVTLEVLREALVGNDSQPLCELLFFELTAIFAABEEEEVAKEBQLTADTKGCSLKS 660
DB 601 GVTLEVLREALVGNDSQPLCELLFFELTAIFAABEEEEVAKEBQLTADTKGCSLKS 660

QY 661 DLDSCTLSEILRLHLTASGADVTSANAKRYQKRGGFDTDDACNELRLSNPSLVKLLSS 720
DB 661 DLDSCTLSEILRLHLTASGADVTSANAKRYQKRGGFDTDDACNELRLSNPSLVKLLSS 720

QY 721 TSVYDLTPEGKMKILHALCGKLLTLVSTRDFTIEDYVDILROAKQFRELKAEQHRKEREE 780
DB 721 TSVYDLTPEGKMKILHALCGKLLTLVSTRDFTIEDYVDILROAKQFRELKAEQHRKEREE 780

QY 781 AAARIRKKEEKLKEQEQMKKEQKEDKEORNSTADISIGEEEREDFDTISSEKQTEQ 840
DB 781 AAARIRKKEEKLKEQEQMKKEQKEDKEORNSTADISIGEEEREDFDTISSEKQTEQ 840

QY 841 KELQDDMFTEDEDDPGSHKRRGRKRGQNGKFEFTRQBIQNCVTVRELLTADEEALKQEH 900
DB 841 KELQDDMFTEDEDDPGSHKRRGRKRGQNGKFEFTRQBIQNCVTVRELLTADEEALKQEH 900
```

```
QY 901 ORKEKELLEKIQSAIACTNIPPLGRDRMYRRYWIFFPSIPGLFIBEDYSGLTEDMLLPSPS 960
DB 901 ORKEKELLEKIQSAIACTNIPPLGRDRMYRRYWIFFPSIPGLFIBEDYSGLTEDMLLPSPS 960

QY 961 SFQNVQSDPQVSTKTGTEPLMSESTSNIDQGRDHSVQLPKPVHKPNRWCFYSCEQLD 1020
DB 961 SFQNVQSDPQVSTKTGTEPLMSESTSNIDQGRDHSVQLPKPVHKPNRWCFYSCEQLD 1020

QY 1021 QLIIBALNSRGHRESALKETLLQEKSRICAQLARFSEKHFSDPKPQDPSKTYSGRGSN 1080
DB 1021 QLIIBALNSRGHRESALKETLLQEKSRICAQLARFSEKHFSDPKPQDPSKTYSGRGSN 1080

QY 1081 AYDPSQCAEQLELRURDLDDIEDRIYQGTLCGAIKVTDRHWRSALESGRYELLSEN 1140
DB 1081 AYDPSQCAEQLELRURDLDDIEDRIYQGTLCGAIKVTDRHWRSALESGRYELLSEN 1140

QY 1141 KENGIITKTVNEDVEEMEIDEQTKVIVKDRLLAGIKTETPSTVSTNASTPQSVSVVHYLAM 1200
DB 1141 KENGIITKTVNEDVEEMEIDEQTKVIVKDRLLAGIKTETPSTVSTNASTPQSVSVVHYLAM 1200

QY 1201 ALFOIEGIERRFLKAPLDASDSGRSYKTVLDRWRESLLSSASLSQVFLHLSLTLDRSVI 1260
DB 1201 ALFOIEGIERRFLKAPLDASDSGRSYKTVLDRWRESLLSSASLSQVFLHLSLTLDRSVI 1260

QY 1261 SKSILNARCKICRKKGAENNVLCGDCGRGHHTYCVRPKLTVPEDGMWFCPECRPKQRCR 1320
DB 1261 SKSILNARCKICRKKGAENNVLCGDCGRGHHTYCVRPKLTVPEDGMWFCPECRPKQRCR 1320

QY 1321 RLSFRQRPSESDVEDSDMGDEDEVDGDEEQSSEEEYEVEQDEDDSQEEBESVLPK 1380
DB 1321 RLSFRQRPSESDVEDSDMGDEDEVDGDEEQSSEEEYEVEQDEDDSQEEBESVLPK 1380

QY 1381 RGRPOVRLPVKTRGKLSLSSFSRQGOQEPGRYPRSQOSTPKTTVSSKTGRSLRKINSAP 1440
DB 1381 RGRPOVRLPVKTRGKLSLSSFSRQGOQEPGRYPRSQOSTPKTTVSSKTGRSLRKINSAP 1440

QY 1441 PTETKSLRIASRSTHSHGLOADVVELLSPRKRGKRSANNTPENSPNPNFRVIAT 1500
DB 1441 PTETKSLRIASRSTHSHGLOADVVELLSPRKRGKRSANNTPENSPNPNFRVIAT 1500

QY 1501 KSSQSRQSVNIASKLSLQESSEKRCRKRQSPSPVTLGRSSRQGRQGVHLSAFEQLV 1560
DB 1501 KSSQSRQSVNIASKLSLQESSEKRCRKRQSPSPVTLGRSSRQGRQGVHLSAFEQLV 1560

QY 1561 VELVRHDDSWPFLKLVSKIQVPDYDIKKPIALNIIREKVNKCEYKLASEFIDDIELMF 1620
DB 1561 VELVRHDDSWPFLKLVSKIQVPDYDIKKPIALNIIREKVNKCEYKLASEFIDDIELMF 1620

QY 1621 SNCFEYNPRNTSEAKAGTRLOAFPHIOAKLGLHVTSPNVDOVSTPPAAKKSRI 1674
DB 1621 SNCFEYNPRNTSEAKAGTRLOAFPHIOAKLGLHVTSPNVDOVSTPPAAKKSRI 1674
```

RESULT 2

```
US-10-376-537-1
; Sequence 1, Application US/10376537
; Publication No. US200302440SA1
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042001
; CURRENT APPLICATION NUMBER: US/10/376,537
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US/09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 73
```

; SOFTWARE: FastSeq for Windows Version 4.0			; SEQ ID NO 1		
; LENGTH: 1674			; TYPE: PRT		
; ORGANISM: Homo sapiens			; ORGANISM: Homo sapiens		
US-10-376-537-1			US-10-376-537-1		
Query Match 100.0%; Score 8631; DB 15; Length 1674;			Query Match 100.0%; Score 8631; DB 15; Length 1674;		
Best Local Similarity 100.0%; Pred. No. 0;			Best Local Similarity 100.0%; Pred. No. 0;		
Matches 1674; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			Matches 1674; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
QY	1	MEDASESRGAPLNNVLPGLSPVSVTVGCKSHRVANKKVEARSEKLLPTALPPSE 60	QY	1021	OLIEALNSRGHRESALKETLLOEKSRICAOQLARFSEKPHFSDKQPDSPKPTYSRGRSSN 1080
DB	1	MEDASESRGAPLNNVLPGLSPVSVTVGCKSHRVANKKVEARSEKLLPTALPPSE 60	DB	1021	OLIEALNSRGHRESALKETLLOEKSRICAOQLARFSEKPHFSDKQPDSPKPTYSRGRSSN 1080
QY	61	PKVDQKLPRSSRRSGGGTQPPARSRAVAAGAAAGAPGERSPLGRRVSPRCLCSG 120	QY	1081	AYDSQMAEAKOLELRLDFLLDIEDRIYQGTDLGAIKVTDRHIWRSALLESYELLSEN 1140
DB	61	PKVDQKLPRSSRRSGGGTQPPARSRAVAAGAAAGAPGERSPLGRRVSPRCLCSG 120	DB	1081	AYDSQMAEAKOLELRLDFLLDIEDRIYQGTDLGAIKVTDRHIWRSALLESYELLSEN 1140
QY	121	EGGVAVGVIAKGRGRGDSRRAPGGREMPLLHRKPFVROKPPADLRPEEVYCKVT 180	QY	1141	KENGIITKTVNEDVEEMEIDEQTKVIVKDRLLGKITETPTSTVSTNASTPQSVSVVHYLAM 1200
DB	121	EGGVAVGVIAKGRGRGDSRRAPGGREMPLLHRKPFVROKPPADLRPEEVYCKVT 180	DB	1141	KENGIITKTVNEDVEEMEIDEQTKVIVKDRLLGKITETPTSTVSTNASTPQSVSVVHYLAM 1200
QY	181	NEIFRHYDDFFERTILCNLSVWSCAVTGRPGLTQYQEALESEKKARONLQSPPEPLIIPVL 240	QY	1201	ALFQIEQGIERRFLKAPLDASDSGRSYKTVLDLRWRESLLSSASLSQVFLHLSTLDRSVITW 1260
DB	181	NEIFRHYDDFFERTILCNLSVWSCAVTGRPGLTQYQEALESEKKARONLQSPPEPLIIPVL 240	DB	1201	ALFQIEQGIERRFLKAPLDASDSGRSYKTVLDLRWRESLLSSASLSQVFLHLSTLDRSVITW 1260
QY	241	YLTSLTHRSRLHEICODIFAYVKDRYFVEETVEVIRNNGARLOCTILEVLPSSHONGFAN 300	QY	1261	SKSILNARCKICRKGDAENMVLCDCDGRGHHTYCVRPKLKTVPEGDMFCPECRPKQRCR 1320
DB	241	YLTSLTHRSRLHEICODIFAYVKDRYFVEETVEVIRNNGARLOCTILEVLPSSHONGFAN 300	DB	1261	SKSILNARCKICRKGDAENMVLCDCDGRGHHTYCVRPKLKTVPEGDMFCPECRPKQRCR 1320
QY	301	GHVNSVDGETIISDSSETQSCFQNGKKDAIDPLLFYKVQPTKKELHESAIVKAT 360	QY	1321	RLSFRQRPLESDEVEDSMGDEDEVDCDEEGSEEEYEVEQDEDDSQEEEVSLPK 1380
DB	301	GHVNSVDGETIISDSSETQSCFQNGKKDAIDPLLFYKVQPTKKELHESAIVKAT 360	DB	1321	RLSFRQRPLESDEVEDSMGDEDEVDCDEEGSEEEYEVEQDEDDSQEEEVSLPK 1380
QY	361	QISREKHLFSRDKLFLKQHCPEQGVITKASSLSYTKIAEQDFSYFFPDPTTFPS 420	QY	1381	RGRQVRLPVKTRGKLSSSFSSRGQQQEPGYPSPRSQQSTPKTTVSSKTGRSLRKINSAP 1440
DB	361	QISREKHLFSRDKLFLKQHCPEQGVITKASSLSYTKIAEQDFSYFFPDPTTFPS 420	DB	1381	RGRQVRLPVKTRGKLSSSFSSRGQQQEPGYPSPRSQQSTPKTTVSSKTGRSLRKINSAP 1440
QY	421	PANRRRGPPKRIHISQEDNVANKOTLASYSKATKRDKLLKQEMKSLAFKAKLARE 480	QY	1441	PTETKSLRIASRSTRSHGHPLOADVVELLSPPRRKCRKSNANTPENSPPNFRVIAT 1500
DB	421	PANRRRGPPKRIHISQEDNVANKOTLASYSKATKRDKLLKQEMKSLAFKAKLARE 480	DB	1441	PTETKSLRIASRSTRSHGHPLOADVVELLSPPRRKCRKSNANTPENSPPNFRVIAT 1500
QY	481	KADALEAKKXKEDKEKREELKKTIVBERLKKKEERLKVBEREKREKREKRYVE 540	QY	1501	KSSQSRSVNTASKLSLQESKRCRCRKRQPEPSPVTLGRRSSQGVHLSAFEQLV 1560
DB	481	KADALEAKKXKEDKEKREELKKTIVBERLKKKEERLKVBEREKREKREKRYVE 540	DB	1501	KSSQSRSVNTASKLSLQESKRCRCRKRQPEPSPVTLGRRSSQGVHLSAFEQLV 1560
QY	541	YLKQWSKPREDMCEDDLKELPEPTPVKTRLPPEIFGDALMWLEFLNATGELFDLQDEFPD 600	QY	1561	VELVRHDSWPFLKLVSKIQVDDYDIKKKPIALNIIREKVNKCEYKLASEFIDDIELMF 1620
DB	541	YLKQWSKPREDMCEDDLKELPEPTPVKTRLPPEIFGDALMWLEFLNATGELFDLQDEFPD 600	DB	1561	VELVRHDSWPFLKLVSKIQVDDYDIKKKPIALNIIREKVNKCEYKLASEFIDDIELMF 1620
QY	601	GVTLEVLBEALVGNDSGFLPCELLFFFLTAIFQALAESEEEVAKBOLTADTKGCSLSKL 660	QY	1621	SNCFEYNPRNTSEAKAGTRLOAFPHIOAKLGLHVTSPSNVDQVSTPPAAKKSRI 1674
DB	601	GVTLEVLBEALVGNDSGFLPCELLFFFLTAIFQALAESEEEVAKBOLTADTKGCSLSKL 660	DB	1621	SNCFEYNPRNTSEAKAGTRLOAFPHIOAKLGLHVTSPSNVDQVSTPPAAKKSRI 1674
QY	661	DLDSCTLSEILRLHLTASGADVTSANAKRYQKRGGFATDDACMELSNPSLVKLLSS 720	RESULT 3		
DB	661	DLDSCTLSEILRLHLTASGADVTSANAKRYQKRGGFATDDACMELSNPSLVKLLSS 720	US-10-702-148-1		
QY	721	TSVYDLTGEKMKILHALCGKLLTLVSTRDFIEDYVDILROAKQFRELKABQHRKEEE 780	; Sequence 1, Application US/10702148		
DB	721	TSVYDLTGEKMKILHALCGKLLTLVSTRDFIEDYVDILROAKQFRELKABQHRKEEE 780	; Publication No. US20040063145A1		
QY	781	AAARIRKKEEKLKEQEKMKKEKQEKLEDEORNSTADISIGEEEREDFTSIESKQTEQ 840	; GENERAL INFORMATION:		
DB	781	AAARIRKKEEKLKEQEKMKKEKQEKLEDEORNSTADISIGEEEREDFTSIESKQTEQ 840	; APPLICANT: Jones, Michael H.		
QY	841	KELQDMFTEDDDPGSHKRRGRGKQNGKFEFTRQEQINCVTRRELLTADEEALKQEH 900	; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR		
DB	841	KELQDMFTEDDDPGSHKRRGRGKQNGKFEFTRQEQINCVTRRELLTADEEALKQEH 900	; FILE REFERENCE: 06501-042002		
QY	901	QRKEKELLEKIQSAIACNIPFLGRDRMYRRIWIPSPGLFIEDYSGLTEDMLLPPRS 960	; CURRENT APPLICATION NUMBER: US/10/702,148		
DB	901	QRKEKELLEKIQSAIACNIPFLGRDRMYRRIWIPSPGLFIEDYSGLTEDMLLPPRS 960	; PRIORITY FILING DATE: 2003-11-05		
			; PRIOR FILING DATE: 2001-04-20		
			; PRIOR APPLICATION NUMBER: US 09/418,710		
			; PRIOR FILING DATE: 1999-10-15		
			; PRIOR APPLICATION NUMBER: PCT/JP98/01783		
			; PRIOR FILING DATE: 1998-04-17		
			; PRIOR APPLICATION NUMBER: JP 9/310027		
			; PRIOR FILING DATE: 1997-10-24		
			; PRIOR APPLICATION NUMBER: JP 9/116570		
			; PRIOR FILING DATE: 1997-04-18		
			; NUMBER OF SEQ ID NOS: 72		
			; SOFTWARE: FastSeq for Windows Version 4.0		

```
; SEQ ID NO 1
; LENGTH: 1674
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-702-148-1

Query Match      100.0%; Score 8631; DB 15; Length 1674;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1674; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MEDASESRGVAPLINNVVLPGSPLSLPVSVTGCKSHRVANKVYKVEARSEKLLPTALPPSE 60
DB 1 MEDASESRGVAPLINNVVLPGSPLSLPVSVTGCKSHRVANKVYKVEARSEKLLPTALPPSE 60

QY 61 PKVQDLPRSSRRSGGGTGPAPARSRAVAGEAAAGAPRGSPGGRVSRPCLCSG 120
DB 61 PKVQDLPRSSRRSGGGTGPAPARSRAVAGEAAAGAPRGSPGGRVSRPCLCSG 120

QY 121 EGGQVAVGVIAKGRGRRGDSRRAPGGREMPLLHRKPFVQKPPADLRPDVEEYFYCKVT 180
DB 121 EGGQVAVGVIAKGRGRRGDSRRAPGGREMPLLHRKPFVQKPPADLRPDVEEYFYCKVT 180

QY 181 NEIFRHYDDFFERTILCNLSVMSCAVTCRPGLTQYQEALESKAKARQLQSPPEPLIIPVL 240
DB 181 NEIFRHYDDFFERTILCNLSVMSCAVTCRPGLTQYQEALESKAKARQLQSPPEPLIIPVL 240

QY 241 YLTSLTHRSRLHEICDDIFAYVKORYFVEETVEVIRNNGARLQCTILEVLPPSHONGFAN 300
DB 241 YLTSLTHRSRLHEICDDIFAYVKORYFVEETVEVIRNNGARLQCTILEVLPPSHONGFAN 300

QY 301 GHVNSVDCGTTIISDSDSETQSCSFQNGKKKDAIDPLLFYKVQPTKKELHESAIVKAT 360
DB 301 GHVNSVDCGTTIISDSDSETQSCSFQNGKKKDAIDPLLFYKVQPTKKELHESAIVKAT 360

QY 361 QISRKKLFSRDKLFLKQHCPEQGVIKKASSLSYTKIAEQDFSYFFPDDPTTFIS 420
DB 361 QISRKKLFSRDKLFLKQHCPEQGVIKKASSLSYTKIAEQDFSYFFPDDPTTFIS 420

QY 421 PANRRGPPKRIHISQEDNVANKOTLASYSKATKEDKLLQKQEMKSLAFKAKLKE 480
DB 421 PANRRGPPKRIHISQEDNVANKOTLASYSKATKEDKLLQKQEMKSLAFKAKLKE 480

QY 481 KADALEAKKKEKEDKKEKREELKKTVEBERLKKKEKERLKVREKEREKLEERKKYVE 540
DB 481 KADALEAKKKEKEDKKEKREELKKTVEBERLKKKEKERLKVREKEREKLEERKKYVE 540

QY 541 YLKQWSKPRDMECDLDELPEPTPVKTRLPPEIFGDALMVLEFLNAGFELFDLQDEPPD 600
DB 541 YLKQWSKPRDMECDLDELPEPTPVKTRLPPEIFGDALMVLEFLNAGFELFDLQDEPPD 600

QY 601 GVTLEVLBEALVGNDSGPLELFFFLTAIFQALAESEEEVAKVKEQLTDADTKGCSLSKL 660
DB 601 GVTLEVLBEALVGNDSGPLELFFFLTAIFQALAESEEEVAKVKEQLTDADTKGCSLSKL 660

QY 661 DLDSCTLSEILRLHLIAGADVTSANAKRYQKRGGFATDDACHNELRLSNPNSLVKKLSS 720
DB 661 DLDSCTLSEILRLHLIAGADVTSANAKRYQKRGGFATDDACHNELRLSNPNSLVKKLSS 720

QY 721 TSVYDLTPGCKMKILHALCGKLLTLVSTRDFTIEDYVDILRQAKQFRELKAEQHRKERE 780
DB 721 TSVYDLTPGCKMKILHALCGKLLTLVSTRDFTIEDYVDILRQAKQFRELKAEQHRKERE 780

QY 781 AAARIRKKEKELKEQEQKMKKEQKLEKEDQBNSTADISIGEEREDFDTSIESKQTEQ 840
DB 781 AAARIRKKEKELKEQEQKMKKEQKLEKEDQBNSTADISIGEEREDFDTSIESKQTEQ 840

QY 841 KELDDMTEDDEDDPGSHKRRGRKRGQNGKFETRQEQINCVTRRELLTADDEEALKEH 900
DB 841 KELDDMTEDDEDDPGSHKRRGRKRGQNGKFETRQEQINCVTRRELLTADDEEALKEH 900

QY 901 QRKEKELLEKIQSAIACTNIPFLGRDRMYRRYWIPIPSIPGLFIEDYSGLTEDMLPRPS 960
DB 901 QRKEKELLEKIQSAIACTNIPFLGRDRMYRRYWIPIPSIPGLFIEDYSGLTEDMLPRPS 960

QY 961 SFQNNVQSDPQVSTKTGTEPLMSESTSNIDQPRDHSVQLPKPVHKPNRMWCFYSCEQLD 1020
DB 961 SFQNNVQSDPQVSTKTGTEPLMSESTSNIDQPRDHSVQLPKPVHKPNRMWCFYSCEQLD 1020

QY 1021 QLIIEALNSRGHRESALKETLLQEKSRICAOQLARFSEKPFHSDKPQDPDKPTYSGRGSN 1080
DB 1021 QLIIEALNSRGHRESALKETLLQEKSRICAOQLARFSEKPFHSDKPQDPDKPTYSGRGSN 1080

QY 1081 AYDSQWCAEQLELRRLRDLFLDIEDRIYOGTLCALIKVTDORHIIWRSALSGRYELLSEN 1140
DB 1081 AYDSQWCAEQLELRRLRDLFLDIEDRIYOGTLCALIKVTDORHIIWRSALSGRYELLSEN 1140

QY 1141 KENGIIKTVAEDVEMEIDEOTKVIKDRLLIGIKTETPTSTVSTNASTPQSQSVSSVHYLAM 1200
DB 1141 KENGIIKTVAEDVEMEIDEOTKVIKDRLLIGIKTETPTSTVSTNASTPQSQSVSSVHYLAM 1200

QY 1201 ALFQIEQIGIERFLKAPLDASDSGRSYKTVLDRWRESLLSSASLSQVFLHLSTLDRSVI 1260
DB 1201 ALFQIEQIGIERFLKAPLDASDSGRSYKTVLDRWRESLLSSASLSQVFLHLSTLDRSVI 1260

QY 1261 SKSTILNARCKICRKKGDAENMVLCDGCDRGHHTYCVRPKLTVPEDGFCFCECRPKQRCR 1320
DB 1261 SKSTILNARCKICRKKGDAENMVLCDGCDRGHHTYCVRPKLTVPEDGFCFCECRPKQRCR 1320

QY 1321 RLSFRQPSLESDEDDVDSMGDEDEVDGDEEQSEEEVEVEQDEDDSQEEBEVSLPK 1380
DB 1321 RLSFRQPSLESDEDDVDSMGDEDEVDGDEEQSEEEVEVEQDEDDSQEEBEVSLPK 1380

QY 1381 RGRPVQLPVKTRCKLSSSFSSRQCGQEPGRYPSPRSQOSTPKTTVSSKTGRSLRKINSAP 1440
DB 1381 RGRPVQLPVKTRCKLSSSFSSRQCGQEPGRYPSPRSQOSTPKTTVSSKTGRSLRKINSAP 1440

QY 1441 PTEKSLRIASRSTRSHSGPLQADVVELLSPRKRGRKSKANNTPENSPNFPNFRVIAT 1500
DB 1441 PTEKSLRIASRSTRSHSGPLQADVVELLSPRKRGRKSKANNTPENSPNFPNFRVIAT 1500

QY 1501 KSSQSSQSVNIASKLSLQESSEKRCRKQSPSPSVTLGRSSRGROGVHLSAFEOLV 1560
DB 1501 KSSQSSQSVNIASKLSLQESSEKRCRKQSPSPSVTLGRSSRGROGVHLSAFEOLV 1560

QY 1561 VELVRHDSPPFLKLVSKIQVPDYDIIRKPIALNIIIRKVNKCEYKLASFIDDIELMF 1620
DB 1561 VELVRHDSPPFLKLVSKIQVPDYDIIRKPIALNIIIRKVNKCEYKLASFIDDIELMF 1620

QY 1621 SNCFEYNPRNTSEAKAGTRLOAFHIOAQKLGHLVTPSNVDQVSTPPAAKKSRI 1674
DB 1621 SNCFEYNPRNTSEAKAGTRLOAFHIOAQKLGHLVTPSNVDQVSTPPAAKKSRI 1674

RESULT 4
US-09-839-479-69
; Sequence 69, Application US/09839479
; Publication No. US2002003979A1
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042002
; CURRENT APPLICATION NUMBER: US/09/839,479
; CURRENT FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 69
; LENGTH: 1673
; TYPE: PRT
```


; ORGANISM: Homo sapiens									
US-09-839-473-69									
Query Match 99.5%; Score 8588.5; DB 9; Length 1673;									
Best Local Similarity 99.6%; Pred. No. 0;									
Matches 1667; Conservative 3; Mismatches 3; Indels 1; Gaps 1;									
Qy	1	MEDAS	SRG	VAP	LNN	VLP	SG	PS	LS
Db	1	MEDAS	SRG	VAP	LNN	VLP	SG	PS	LS
Qy	61	PKVD	OKL	PRSS	RRG	GGT	Q	PA	RS
Db	61	PKVD	OKL	PRSS	RRG	GGT	Q	PA	RS
Qy	121	EGG	Q	VAV	G	V	I	A	G
Db	121	EGG	Q	VAV	G	V	I	A	G
Qy	181	NEI	F	R	H	D	D	F	F
Db	181	NEI	F	R	H	D	D	F	F
Qy	241	YLT	S	L	T	H	R	S	L
Db	241	YLT	S	L	T	H	R	S	L
Qy	301	GHV	N	S	V	D	G	E	T
Db	301	GHV	N	S	V	D	G	E	T
Qy	361	QIS	R	K	H	L	F	S	R
Db	361	QIS	R	K	H	L	F	S	R
Qy	421	PAN	R	R	G	P	P	K	I
Db	421	PAN	R	R	G	P	P	K	I
Qy	481	KAD	A	E	A	K	K	E	D
Db	481	KAD	A	E	A	K	K	E	D
Qy	541	Y	L	K	O	N	S	K	P
Db	541	Y	L	K	O	N	S	K	P
Qy	601	G	V	T	L	E	A	L	V
Db	600	G	V	T	L	E	A	L	V
Qy	661	D	L	D	S	C	T	L	S
Db	660	D	L	D	S	C	T	L	S
Qy	721	T	S	V	D	L	T	P	G
Db	720	T	S	V	D	L	T	P	G
Qy	781	A	A	A	R	K	E	E	K
Db	780	A	A	A	R	K	E	E	K
Qy	841	K	E	L	D	O	M	F	T
Db	840	K	E	L	D	O	M	F	T
Qy	901	O	R	K	E	L	E	K	I
Db	900	O	R	K	E	L	E	K	I
Qy	961	S	F	O	N	N	V	Q	S
Db	960	S	F	O	N	N	V	Q	S

Db	960	S	F	O	N	N	V	Q	S
Qy	1021	Q	L	I	E	A	L	N	S
Db	1020	Q	L	I	E	A	L	N	S
Qy	1081	A	Y	D	P	S	O	M	C
Db	1080	A	Y	D	P	S	O	M	C
Qy	1141	K	E	N	G	I	I	K	T
Db	1140	K	E	N	G	I	I	K	T
Qy	1201	A	L	F	I	E	O	G	I
Db	1200	A	L	F	I	E	O	G	I
Qy	1261	S	K	S	I	L	N	A	R
Db	1260	S	K	S	I	L	N	A	R
Qy	1321	R	L	S	F	R	O	R	P
Db	1320	R	L	S	F	R	O	R	P
Qy	1381	R	G	R	P	O	V	R	L
Db	1380	R	G	R	P	O	V	R	L
Qy	1441	P	T	E	T	K	S	L	R
Db	1440	P	T	E	T	K	S	L	R
Qy	1501	K	S	S	O	S	R	V	N
Db	1500	K	S	S	O	S	R	V	N
Qy	1561	V	E	L	V	H	D	S	W
Db	1560	V	E	L	V	H	D	S	W
Qy	1621	S	N	C	F	E	Y	N	P
Db	1620	S	N	C	F	E	Y	N	P

RESULT 5

US-10-376-537-70
; Sequence 70, Application US/10376537
; Publication No. US20030224405A1
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042001
; CURRENT APPLICATION NUMBER: US/10/376,537
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US/09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 70
; LENGTH: 1673
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-376-537-70

Query Match				99.5%;	Score 8588.5;	DB 15;	Length 1673;
Best Local Similarity				99.6%;	Pred. No. 0;		
Matches 1667;				Conservative	3;	Mismatches	3;
						Indels	1;
						Gaps	1;
Qy	1	MEDASESRGVAPLINNVVLPGSLPLSVSTGCKSHRVANKVVEARSEKLLPTALPPSE	60				
Db	1	MEDASESRGVAPLINNVVLPGSLPLSVSTGCKSHRVANKVVEARSEKLLPTALPPSE	60				
Qy	61	PKVQDKLPRSSERGGSGGTGTPPARSRAVAAGEAAAGAPERGSPGLGRVSRPRLCSG	120				
Db	61	PKVQDKLPRSSERGGSGGTGTPPARSRAVAAGEAAAGAPERGSPGLGRVSRPRLCSG	120				
Qy	121	EGGVAVGVIAGKRRGRDGRSRRAPGRENPLLRKPFVVRQPPADLRPDEEVFYCKVT	180				
Db	121	EGGVAVGVIAGKRRGRDGRSRRAPGRENPLLRKPFVVRQPPADLRPDEEVFYCKVT	180				
Qy	181	NEIFRHYDDPFERTILCNLSVWCNAVTPORGLTYQEALESEKKARQNLQSPPEBLIIPVL	240				
Db	181	NEIFRHYDDPFERTILCNLSVWCNAVTPORGLTYQEALESEKKARQNLQSPPEBLIIPVL	240				
Qy	241	YLTSI,THRSRLHEICDDIFAVVKDRIYFVEETVEVIRNNGARLQCTILEVLPPSHONGPAN	300				
Db	241	YLTSI,THRSRLHEICDDIFAVVKDRIYFVEETVEVIRNNGARLQCTILEVLPPSHONGPAN	300				
Qy	301	GHVNSVDGETIIISDSDDSETQSCSFQNGKKDAIDPLLFYKYQPTKKELHESAIVKAT	360				
Db	301	GHVNSVDGETIIISDSDDSETQSCSFQNGKKDAIDPLLFYKYQPTKKELHESAIVKAT	360				
Qy	361	QISRRKHLFSRDKULFLKQHCPEQGVIKIKASSLSYTKIABODFSYFFPDPTFTFS	420				
Db	361	QISRRKHLFSRDKULFLKQHCPEQGVIKIKASSLSYTKIABODFSYFFPDPTFTFS	420				
Qy	421	PANRRRGPXRIIHSQDNVANKOTLASYSKATKEDKLLKOEEMKSLAFKAKLRE	480				
Db	421	PANRRRGPXRIIHSQDNVANKOTLASYSKATKEDKLLKOEEMKSLAFKAKLRE	480				
Qy	481	KADALEAKKKEKEDKKEERLKVIVERERLKKKEERLKVVEREKERLREERKYVE	540				
Db	481	KADALEAKKKEKEDKKEERLKVIVERERLKKKEERLKVVEREKERLREERKYVE	540				
Qy	541	YLKQWSKPRDEMCDDLKELPEPTPVKTRLPPEIFGDALMVLEFLNAGELFDLQDEPPD	600				
Db	541	Y-KQWSKPRDEMCDDLKELPEPTPVKTRLPPEIFGDALMVLEFLNAGELFDLQDEPPD	599				
Qy	601	GVTLEVLBEALVGNDSQPLCELLPFFLTALFOAIAEBEEVEAKEQLTDADTKGCSLSL	660				
Db	600	GVTLEVLBEALVGNDSQPLCELLPFFLTALFOAIAEBEEVEAKEQLTDADTKGCSLSL	659				
Qy	661	DLDSCTLSEILRLHLILASGADVTSANAKYRQKRGGFATDDACMELRSLNPSLVKKLSS	720				
Db	660	DLDSCTLSEILRLHLILASGADVTSANAKYRQKRGGFATDDACMELRSLNPSLVKKLSS	719				
Qy	721	TSVYDLTPGERKWKILHALCGKLLTLVSTRDFIEDYVDILROAKQEFRELKAEQHRKEREE	780				
Db	720	TSVYDLTPGERKWKILHALCGKLLTLVSTRDFIEDYVDILROAKQEFRELKAEQHRKEREE	779				
Qy	781	AAARIRKKEKELKEQEQMKKEQKLEDEQRNSTADISICEEREDFDTSIESKOTEQ	840				
Db	780	AAARIRKKEKELKEQEQMKKEQKLEDEQRNSTADISICEEREDFDTSIESKOTEQ	839				
Qy	841	KELODMTEDEDDPGSHKRRGRGQNGKFEPTROEQINCVTRELLTADBEELAKQEH	900				
Db	840	KELODMTEDEDDPGSHKRRGRGQNGKFEPTROEQINCVTRELLTADBEELAKQEH	899				
Qy	901	QRKEKELLEKIOSAIACTNIFPLGRDRMYRRIYFPSPIGLFIEDYSGLTEDMLLPPRS	960				
Db	900	QRKEKELLEKIOSAIACTNIFPLGRDRMYRRIYFPSPIGLFIEDYSGLTEDMLLPPRS	959				
Qy	961	SFQNNVQSQDQVSTKTGTEPLMSESTSNIDQPRDHSVOLPKPVHKPNRWCIFYSSCEQLD	1020				
Db	960	SFQNNVQSQDQVSTKTGTEPLMSESTSNIDQPRDHSVOLPKPVHKPNRWCIFYSSCEQLD	1019				
Qy	1021	QLIEALNSRGHRESALKETTLQKSRICAQLARFSEKHFSDKQPQDSKPTYSRGRSSN	1080				

Db	1020	QLIEALNSRGHRESALKETTLQKSRICAQLARFSEKHFSDKQPQDSKPTYSRGRSSN	1079				
Qy	1081	AYDSQWCAEQELRLRDLFDIEDRIYOGTILGAIKVTDORHIWRSALSGRYELLSEEN	1140				
Db	1080	AYDSQWCAEQELRLRDLFDIEDRIYOGTILGAIKVTDORHIWRSALSGRYELLSEEN	1139				
Qy	1141	KENGIIKTVNEDVEEMIDEOTKIVKDRLIGIKTETPSTVSTNASTPQSVSSVVHYLAM	1200				
Db	1140	KENGIIKTVNEDVEEMIDEOTKIVKDRLLGIKTETPSTVSTNASTPQSVSSVVHYLAM	1199				
Qy	1201	ALFOIEGIERFLKAPLDASDGRSYKTVLDRWRESLLSSASISQVFLHJLSTLDRSVI	1260				
Db	1200	ALFOIEQGLERRFLKAPLDASDGRSYKTVLDRWRESLLSSASISQVFLHJLSTLDRSVI	1259				
Qy	1261	SKSTLNARCKICRKKGAENNVLCDCGDRGHHTYCVRPKLTVPPEGWFCPECRPKQCR	1320				
Db	1260	SKSTLNARCKICRKKGAENNVLCDCGDRGHHTYCVRPKLTVPPEGWFCPECRPKQCR	1319				
Qy	1321	RLSFRQPSLESDEVEDSMGGEDEVDGDEEQSEEEYEVEQDEDDDSQEEBEVSLPK	1380				
Db	1320	RLSFRQPSLESDEVEDSMGGEDEVDGDEEQSEEEYEVEQDEDDDSQEEBEVSLPK	1379				
Qy	1381	RGRPOVLPVKTRKLSSSFSSSRGQOQEPGRYPSPRSQOSTPCTTVSSKTGRSLRKINSAP	1440				
Db	1380	RGRPOVLPVKTRKLSSSFSSSRGQOQEPGRYPSPRSQOSTPCTTVSSKTGRSLRKINSAP	1439				
Qy	1441	PTETKSIRIASRSTRHSHGLOADVVELLSPRKRGRKSANNTPENSPNPNFRVIAT	1500				
Db	1440	PTETKSIRIASRSTRHSHGLOADVVELLSPRKRGRKSANNTPENSPNPNFRVIAT	1499				
Qy	1501	KSSQSSSVNIASKLSLQESKRRCKRQSPSPSVTLGRSSRGQGVHLSAFEOLV	1560				
Db	1500	KSSQSSSVNIASKLSLQESKRRCKRQSPSPSVTLGRSSRGQGVHLSAFEOLV	1559				
Qy	1561	VELVRHDSWPFLKLVSKIQVPDYDIIKKPIALNIIIREKVNKCEYKLASFIDDIELMP	1620				
Db	1560	VELVRHDSWPFLKLVSKIQVPDYDIIKKPIALNIIIREKVNKCEYKLASFIDDIELMP	1619				
Qy	1621	SNCFEYNPRNTSEAKAGTRLOAFPHIOAKLGLHVTPTSNVDQVSTPPAAKKSRI	1674				
Db	1620	SNCFEYNPRNTSEAKAGTRLOAFPHIOAKLGLHVTPTSNVDQVSTPPAAKKSRI	1673				
RESULT 6							
US-10-702-148-69							
; Sequence 69, Application US/10702148							
; Publication No. US20040063145A1							
; GENERAL INFORMATION:							
; APPLICANT: Jones, Michael H.							
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR							
; FILE REFERENCE: 06501-042002							
; CURRENT APPLICATION NUMBER: US/10702,148							
; PRIOR FILING DATE: 2003-11-05							
; PRIOR APPLICATION NUMBER: PRIOR APPLICATION NUMBER: US/09/839,479							
; PRIOR FILING DATE: 2001-04-20							
; PRIOR APPLICATION NUMBER: US 09/418,710							
; PRIOR FILING DATE: 1999-10-15							
; PRIOR APPLICATION NUMBER: PCT/JP98/01783							
; PRIOR FILING DATE: 1998-04-17							
; PRIOR APPLICATION NUMBER: JP 9/310027							
; PRIOR FILING DATE: 1997-10-24							
; PRIOR APPLICATION NUMBER: JP 9/116570							
; PRIOR FILING DATE: 1997-04-18							
; NUMBER OF SEQ ID NOS: 72							
; SOFTWARE: FastSeq for Windows Version 4.0							
; SEQ ID NO 69							
; LENGTH: 1673							
; TYPE: PRT							
; ORGANISM: Homo sapiens							
US-10-702-148-69							
Query Match				99.5%;	Score 8588.5;	DB 15;	Length 1673;

Best Local Similarity 99.6%; Pred. No. 0;			
Matches 1667; Conservative 3; Mismatches 3; Indels 1; Gaps 1;			
QY	1	MEDASESRGVAPLINNVVLPGLSPVSTGCKSHRVANKVVEARSEKLLPTALPPSE	60
Db	1	MEDASESRGVAPLINNVVLPGLSPVSTGCKSHRVANKVVEARSEKLLPTALPPSE	60
QY	61	PKVDQKLPRSSRRSGGTQPPARSRAVAGEAAAGAPERGSPIGRRVSPRCLCSG	120
Db	61	PKVDQKLPRSSRRSGGTQPPARSRAVAGEAAAGAPERGSPIGRRVSPRCLCSG	120
QY	121	EGGVAVGVIAKGRRRGDSRRAPGGREMLLRKPPVRQKPPADLRPDEEVEYCKVT	180
Db	121	EGGVAVGVIAKGRRRGDSRRAPGGREMLLRKPPVRQKPPADLRPDEEVEYCKVT	180
QY	181	NEIFRYDDFFERTILCNLSWSCAVTGRPGITYQEALESEKAKQNLSQFPPELIIPVL	240
Db	181	NEIFRYDDFFERTILCNLSWSCAVTGRPGITYQEALESEKAKQNLSQFPPELIIPVL	240
QY	241	YLTSLTHSRRLHEICDDIPAYVKORYFVEETVEVIRNNGARLQCTILEVLPSSHONGPAN	300
Db	241	YLTSLTHSRRLHEICDDIPAYVKORYFVEETVEVIRNNGARLQCTILEVLPSSHONGPAN	300
QY	301	GHVNSVDGETIIISDSDDSETQSCFONGKKDAIDPLLFKYKVOPTKKELHESAIVKAT	360
Db	301	GHVNSVDGETIIISDSDDSETQSCFONGKKDAIDPLLFKYKVOPTKKELHESAIVKAT	360
QY	361	QISRRKHLFSRDKLFLKQHCPEQGVIKIKASSLSTYKIAEQDFSYFFPDPTTFPS	420
Db	361	QISRRKHLFSRDKLFLKQHCPEQGVIKIKASSLSTYKIAEQDFSYFFPDPTTFPS	420
QY	421	PANRRGPPPKRIHISQEDNVANKOTLASYSKATKRDKLLKQEMKSLAFKAKLKRE	480
Db	421	PANRRGPPPKRIHISQEDNVANKOTLASYSKATKRDKLLKQEMKSLAFKAKLKRE	480
QY	481	KADALEAKKEDKKEKREELKKTVEBERLKKKEKERLKYVEREKEREKLEEKRYVE	540
Db	481	KADALEAKKEDKKEKREELKKTVEBERLKKKEKERLKYVEREKEREKLEEKRYVE	540
QY	541	YLKQWSKPREMDECDLDELPEPTPVKTRLPPEIFGDALMWLEFNALFGEFLDQDEFPD	600
Db	541	Y-KQWSKPREMDECDLDELPEPTPVKTRLPPEIFGDALMWLEFNALFGEFLDQDEFPD	599
QY	601	GVTLEVLBEALVGNDSGLPCELLFFLTALFOAIAEBEEVEAKBQLTADTKGCSLKS	660
Db	600	GVTLEVLBEALVGNDSGLPCELLFFLTALFOAIAEBEEVEAKBQLTADTKGCSLKS	659
QY	661	DLDSCTLSEILRLHLASGADVTSANAKYRQKRGGFATDDACMELRLSNPVLKYLSS	720
Db	660	DLDSCTLSEILRLHLASGADVTSANAKYRQKRGGFATDDACMELRLSNPVLKYLSS	719
QY	721	TSVYDLTPGEKMKILHALCGKLLTLVSTRDFTEDYVDILRQAKQBFRELKAEQHRKERE	780
Db	720	TSVYDLTPGEKMKILHALCGKLLTLVSTRDFTEDYVDILRQAKQBFRELKAEQHRKERE	779
QY	781	AAARIRKKEEKLKEQEKMKKEQBLKEDQERNSTADISIGEEEREDFTSIESKQTEQ	840
Db	780	AAARIRKKEEKLKEQEKMKKEQBLKEDQERNSTADISIGEEEREDFTSIESKQTEQ	839
QY	841	KELDOMTFEDDDPGSHKRRGRKGQNGFKFETRQEQINCVTRELLTADAEELAKQEH	900
Db	840	KELDOMTFEDDDPGSHKRRGRKGQNGFKFETRQEQINCVTRELLTADAEELAKQEH	899
QY	901	QRKEKELLEKQSACTACTNIFPLGRDRMYRRWIIFPSIPGLFIEDYSGLTEDMLPRPS	960
Db	900	QRKEKELLEKQSACTACTNIFPLGRDRMYRRWIIFPSIPGLFIEDYSGLTEDMLPRPS	959
QY	961	SFQNNVQSQDPQVSTKTGEPLMSESTSNIDQPRDHSVQLPKPVHKPNRWCFYSSCEQLD	1020
Db	960	SFQNNVQSQDPQVSTKTGEPLMSESTSNIDQPRDHSVQLPKPVHKPNRWCFYSSCEQLD	1019
QY	1021	QLIEALNSRGHRESALKETLLOEKSRICAQLARFSEKPFHSDKQPKDPTYSRGRSN	1080

Db	1020	QLIEALNSRGHRESALKETLLOEKSRICAQLARFSEKPFHSDKQPKDPTYSRGRSN	1079
QY	1081	AYDSQMAEQLELRDLFLDIEDRIYQGTGLGAIKVTDRHIMRSALSCRYELLSEEN	1140
Db	1080	AYDSQMAEQLELRDLFLDIEDRIYQGTGLGAIKVTDRHIMRSALSCRYELLSEEN	1139
QY	1141	KENGIITKTVNEDVEMEIDEQTKIVIKDRILGIKTETSTVSTNASTPQSQSSVVHYLAM	1200
Db	1140	KENGIITKTVNEDVEMEIDEQTKIVIKDRILGIKTETSTVSTNASTPQSQSSVVHYLAM	1199
QY	1201	ALFOEQGIERRFLKAPLDASDGRSYKTVLDRWRESLSSASLSQVFLHLSTLDRSVIM	1260
Db	1200	ALFOEQGLERRFLKAPLDASDGRSYKTVLDRWRESLSSASLSQVFLHLSTLDRSVIM	1259
QY	1261	SKSILNARCKTCKKGDANMVLCDGCDRGHHTYCVRPKLKTVPEGDWFCPECRPKQCR	1320
Db	1260	SKSILNARCKTCKKGDANMVLCDGCDRGHHTYCVRPKLKTVPEGDWFCPECRPKQCR	1319
QY	1321	RLSRFRPSLESDEVEDSMGDEDDVDGDEEGQSEEEVEVEQEDDSDQEEERVSPLK	1380
Db	1320	RLSRFRPSLESDEVEDSMGDEDDVDGDEEGQSEEEVEVEQEDDSDQEEERVSPLK	1379
QY	1381	RGRQVRLPVKTRGKLSSESSRGQQQPGRYPSRSQQSTPKTTVSSKTGRSLRKINAP	1440
Db	1380	RGRQVRLPVKTRGKLSSESSRGQQQPGRYPSRSQQSTPKTTVSSKTGRSLRKINAP	1439
QY	1441	PTETKSLRIASRSTRHSHGPIQADVFVLLSPRRKRRGRKSNANTPENSPPNFRVIAT	1500
Db	1440	PTETKSLRIASRSTRHSHGPIQADVFVLLSPRRKRRGRKSNANTPENSPPNFRVIAT	1499
QY	1501	KSSQSGRSVNTASKLSLOESKSKRCRCRKQSPSPSVTLGRRSSRGQGVHLSAFEQLV	1560
Db	1500	KSSQSGRSVNTASKLSLOESKSKRCRCRKQSPSPSVTLGRRSSRGQGVHLSAFEQLV	1559
QY	1561	VELVRHDDSWFLKLVSKIQVDDYDIITKPIALNIIREKVNKCYKLASFIDDIELMF	1620
Db	1560	VELVRHDDSWFLKLVSKIQVDDYDIITKPIALNIIREKVNKCYKLASFIDDIELMF	1619
QY	1621	SNCFEYNPRNTSEAKAGTRLOAFFHIOAQKLGHLVTPSNVDQVSTPPPAKKSRI	1674
Db	1620	SNCFEYNPRNTSEAKAGTRLOAFFHIOAQKLGHLVTPSNVDQVSTPPPAKKSRI	1673
RESULT 7			
US-10-264-237-2629			
; Sequence 2629, Application US/10264237			
; Publication No. US20040009491A1			
; GENERAL INFORMATION:			
; APPLICANT: Birse et al.			
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies			
; FILE REFERENCE: P131P1			
; CURRENT APPLICATION NUMBER: US/10/264,237			
; PRIOR FILING DATE: 2002-10-04			
; PRIOR APPLICATION NUMBER: PCT/US01/16450			
; PRIOR FILING DATE: 2001-05-18			
; PRIOR APPLICATION NUMBER: US 60/205,515			
; PRIOR FILING DATE: 2000-05-19			
; NUMBER OF SEQ ID NOS: 2876			
; SOFTWARE: PatentIn Ver. 3.1			
; SEQ ID NO 2629			
; LENGTH: 279			
; TYPE: PRT			
; ORGANISM: Homo sapiens			
US-10-264-237-2629			
Query Match 14.1%; Score 1215; DB 15; Length 279;			
Best Local Similarity 88.4%; Pred. No. 6e-60;			
Matches 245; Conservative 0; Mismatches 0; Indels 32; Gaps 1;			
QY	552	MECDLDELPEPTPVKTRLPPEIFGDALMWLEFNALFGEFLDQDEFPDGVTLVLEAL	611
Db	1	MECDLDELPEPTPVKTRLPPEIFGDALMWLEFNALFGEFLDQDEFPDGVTLVLEAL	60


```
QY 592 FDLQDFPDGVTVLEVEALVNDSEGPICELLFFFLTAIFAIAEBEEVEAKEQJTDAD 651
Db 61 FDLQDFPDGVTVLEVEALVNDSEGPICELLFFFLTAIFAIAEBEEVEAKEQJTDAD 120
QY 652 TKGCSLKSJLDLSCITLRLHILASGADVTSANAKRYQKRGGFDDTDDACMEIRLSN 711
Db 121 TKGCSLKSJLDLSCITLRLHILASGADVTSANAKRYQKRGGFDDTDDACMEIRLSN 180
QY 712 PSLVKLSSTSVYDLTPGKMKILHALCGKL 742
Db 181 PSLVKLSSTSVYDLTPGKMKILHALCGKL 211

RESULT 11
US-09-839-479-29
; Sequence 29, Application US/09839479
; Publication No. US20020039779A1
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042002
; CURRENT APPLICATION NUMBER: US/09/839,479
; CURRENT FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 29
; LENGTH: 1531.
; TYPE: PR1
; ORGANISM: Homo sapiens
US-09-839-479-29

Query Match 10.8%; Score 930.5; DB 9; Length 1531;
Best Local Similarity 23.2%; Pred. No. 3.9e-43;
Matches 399; Conservative 263; Mismatches 526; Indels 529; Gaps 68;

QY 152 PLLHKPFVQKPPADLPDEEVFY-CKVTNIFRHYDDFFERTILCNLSVMSCAVTGRP 210
Db 3 PLLGRKPF----PLVNLPGEEPFTIPIHTQEAFTREYEARELERYSERIWTCKSGS 58
QY 211 GLTYQEALESEKARQNL-QSPF---EPLIIPVLYLTSLTHRSRLHBCDDIFAYVDORY 266
Db 59 QLTHKEAWEEOEVAELKEEPAPWYKLVLEMVH----HNTASLEKLVDTAWLEIMTKY 114
QY 267 FVEETVEVIRNNGARLOCTILEVLP-----PS-----HONGFANGHV 303
Db 115 AVGEECDPEVGKRWMLVKVIKHPLEKVEATEKESKDGACDSSPSKSNSSQIAQDHQ 174
QY 304 NSVDGETIISDSDSETQS-----CSFQNGKKKDAIDPLL-FKKVQPTKXEL 351
Db 175 KK---ETVVKEDEGRESINDPARRSPKLPISLKKGERKWAAPPKFLPKYDV----KLQ 227
QY 352 HESAI---VKATQISRRKHLFSRDKLKLFLKQHC-----EPQEGVIK---IKASSLSTY 399
Db 228 NEDKIISNVPADSLIRTPPNPKVIVFYIRHNALRAGTENAPWVVEDELVKYSLPS- 286
QY 400 KIAEQDFSYFPDPTTIFSPANRRG-----RPPKRIHISQDN----- 440
Db 287 ----KFSDFLDPKYMTLNPSTKRNKNGSPDRKPSK---SKTDNSLSLSPNPKLWC 338
QY 441 -VANKOTLA-----SVRSKATKER----- 458
Db 339 HVHLKKSLSGSPLVKNKSNKSPPEHLEEMKMKMSPNKLHTNFHHPKGGPPAKPGKHS 398
QY 459 DKLLK-----QEMKSLAFKAKLKR----- 479
```

```
Db 399 DKPLKAKGRSKGILNGQXSTGNSKSPKGLKTPKTKMKQMTLLDMAKGTQKMTAPRNSG 458
QY 480 -----SKADALEAKKEKEDKEKR-----EEL 502
Db 459 GTPRTSSKPKHKLPPAALHLIAYKENKDREKRSALSVCISKARLLSSSEDRARLPSEL 518
QY 503 KKIIVEE-----RLKKKEKERLKVEREKERELRE--EKRYKVEYLKQWMSKPR 549
Db 519 RSLVQRYVELLEHKRWASMSBEQRYLKKKREELKKLKEKAKERREKEMLEKQK 578
QY 550 --EDMECDDDLKELPEPTPVKT--RLPPEIFGDALMVLEFLNAFGLFDLQDFPDGVTVLE 605
Db 579 RYEDQELTG-KNLPAFRLVDTPEGLPNTLFGDVAMVVFSLSCYSGLLLPDAQYP--ITAV 635
QY 606 VLEELAVGNDSEG--PLCELLFFFLTAIFAIAEBEEVEAKEQLTADTKGCSLSKSLDLD 663
Db 636 SLMEAL-SADKGGFLYLNRLVILLQTLQTL--QDEIAE----DYGELGKMLSEIPLT 688
QY 664 SCTLSEILRLHILASGADVTSANAKRYQKRGGFDDTDDACMELRLSNPSLVKLSSTSV 723
Db 689 LHSVSELVRLCLRRSDVQEESESGTD-----DNKDSAAFEDNEVDQEFLEKLETSEF 741
QY 724 YDLTPGKMKILHALCGKLTLVSTRDFIEDYVDILROAKQEFRELKAEQHRKEBEAAA 783
Db 742 FELTSEKQLQILTALCHRLMTYSVD-----HMETRQOMSA 778
QY 784 RIRKKEBEKLAKEOEKMEKOEKLEKDEORNSTADISIGEEREDFDTSIESKDEOK-- 841
Db 779 ELWKERLAVLEENDKGAERKQKEMEAKN-----KENGKVENGLGKTDKRRKIV 829
QY 842 ELDDQMFTEDDDPGSHKRGK----RGKRGQNGFKFETRQEQINCVTRELITADBEELK 897
Db 830 KFEPQVDTAEADMISAVKSRLLAIAQK-----KEREIQE-----REMKVKLERQA-E 876
QY 898 QEHORKEKELEK-IQSAIACNTNIF----PLGRDMYRYWIF-PSIPGLFTEEDYSGLT 951
Db 877 EERIRKHAABAKAPQEGIAKAKLVMMRTPIGTDRNHNRYWLFSDDEVFGLFIEK---GWV 933
QY 952 EDMLLPSPSSFQNNVQSDPOVSTKTGP-----LMSESTSNIDQG-PRDHSVOL 1000
Db 934 HDSI-----DVRFNHCKDHTVS---GDIEDYCPRSKKANLGNKSNMTQHGTAATEVAVBT 985
QY 1001 PKPVHKPNRWCFYSSCEQLDQILIEALNSRGHRESALKETLLQEKSRICAQLARFSE--BK 1058
Db 986 TTPKQGNLWFLCDSQKELDELLNCLHPQGRESQLKERL--EK-----RYQDIHS 1035
QY 1059 FHFDKQPDPSKPTYRGRSSNAVDPQSMCAEKQLELRDLRDLIEDIRIYOGTLGAIKV 1118
Db 1036 IHLARKPNLGLK-----SCDGNQ-----ELLNFLRSLDIEVATRLQKGLGYVEE 1080
QY 1119 TDRHWRSALESGRVELLSEENKENGIIKTVNEDEVEEIDEQTKVIVK--DRLLAGIKTE 1176
Db 1081 T-----SEPEA---RVISLEKL-----KQFGCVIALQASVIKKFLQGFMAPKQK 1122
QY 1177 TPTSTVSTNASTPQSVSSVVHYLAMALFOIEQIERRFLKAPLDASDSGRSVKTVLDRWE 1236
Db 1123 RRKLQESDSAKTEEVE-----EKNVVEEAKVAS-----ALEKWK 1158
QY 1237 SLLSASLSQVFLHLSTLDRSVIWSKSLNARCKICRKKGAENMVLCDCGRGHHTYCV 1296
Db 1159 AIRBAQTFSRMHVLLGLMDACIKWMSAENARCKVCPKKGEDDKLILCDECNKAPHLFCL 1218
QY 1297 RPKLKTYPEGDFWCFECPKORCRLSPRQPSLESDVEDSDMGCEDEVDGDEEEQOS 1356
Db 1219 RPALYEVFDGQWQPCAPQATARNRNRNRYTEESASEDESEDESEEE-----BEEBEE 1274
QY 1357 EEEVEVEQDEDDSQEEBEVSLPKRGRPOVRL-PVKT-RGKLSSSFSGRQOOEGRYPS 1414
Db 1275 EEEYEV-----AGLRLRPKTIRGKHSVIPPAAERSGRRRGKPKH 1314
QY 1415 RSQOSTPKTVSSKTRSLRKINSAPPTETKSLRIASRSTRHSHGLOADVFEVLLSPRR 1474
Db 1415 RSQOSTPKTVSSKTRSLRKINSAPPTETKSLRIASRSTRHSHGLOADVFEVLLSPRR 1474
```


Db 1315 STRSQPK-----APPVD-----DAEDEL----- 1334
Qy 1475 KRRGRKSANTPENSFPNFRVIATKSSQSRSVNIASKLSQESSEKRCRKRQSP 1534
Db 1335 -----VJQTKSSRRQSL----- 1347
Qy 1535 SPVTILGRSSRGQGVHLSAFQELVVLVRHDSWPFLKLVSKIQVDDYDIKKPIAL 1594
Db 1348 -----ELQKCEELHKKVYRFSWPPREPVRDEABDYDVITHPMD 1390
Qy 1595 NIIEKVNKCYKLADEFIDDIEMFNSCFYNPNT 1631
Db 1391 QTVQNKCSGYSRVSQEPFLDMKQVFTNAEYVNCGRS 1427

RESULT 12
US-10-376-537-29
; Sequence 29, Application US/10376537
; Publication No. US20030224405A1
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042001
; CURRENT APPLICATION NUMBER: US/10/376,537
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US/09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 29
; LENGTH: 1531
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-376-537-29

Query Match 10.8%; Score 930.5; DB 15; Length 1531;
Best Local Similarity 23.2%; Pred. No. 3.9e-43;
Matches 399; Conservative 263; Mismatches 526; Indels 529; Gaps 68;

Qy 152 PLLHRKPVROKPADLRPDDEVFY-CKVTNEIFRHYDDPFERTILCNLSWSCAVTGRP 210
Db 3 PLLGRKPP-----PLVNPPLGEBEPFTIPIHTQEAFTREBYEARLERYSERIWTCKSTGSS 58

Qy 211 GLTYOEALSEKKAQNL-QSFP---EPLIIPVLYLTSLTHRSRLHEICDDIFAVVKDXY 266
Db 59 QLTHKEAEEBQEAELKEEFPWAYEKLIVEMVH-----HNTASLEKLVDTAWLEIMTKY 114

Qy 267 FVEETVEIRNNGARLOCTILEVLP-----PS-----HQNGFANGHV 303
Db 115 AVGECDFEVGKXMKLVKVIKIHPLEKVBDEATEKSDGACDSPSSDKENSSQIAQDHQ 174

Qy 304 NSVGETIIISDSDSETQS-----CSFGNGKKKDAIDPLL-FKYKVOPTKKEL 351
Db 175 KK---ETVVKEDGRRESINDRARRSPRKLPTSLKKGRKWWAPPKFLPHKYDV---KLQ 227

Qy 352 HESAI---VKATOISRRKHLFSRDKLKLFLKQHC-----EPQEGVIK---IKASSLSTY 399
Db 228 NEDKIISNVPADSLIRTPNKLEIVYFIRHNAIRAGTENAPMVVEDELVKYKYSUPS- 286

Qy 400 KIAEQDFYFPDPPPTFIISPANRRG-----RPPKRIHISOEDN----- 440
Db 287 -----KFSDFLLDPKYMTLNPSTKTKNGTSGPDRKPSKK---SKTDNSLSPLNPKLWC 338

Qy 441 -VANKQTILA-----SYRSKATKER----- 458
Db 339 HVHLKKSLSGLPKVKNKSKSPHEHLEEMKMKMSPNKLHTNFHIPPKGPPAKKPGKHS 398

Qy 459 DKLLK-----OEMKSLAFKAKLKR----- 479
Db 399 DXPLKAKRSGKILNGKSTGNSKSPKGLKTPTKMKQMTLLDMWAKGTQKWTAPRNSG 458

Qy 480 -----EKADALEAKKKEKEDKKR-----BEL 502
Db 459 GTPRTSSKPHKHLPPAALHLIAYYKENKDRKDSALS CVISKTARLLSSSDRALRPEL 518

Qy 503 KKIIVEEB-----RLKKKEERLKVVEREKREKLE---BKRYVYLKQWSKPR 549
Db 519 RSLVQKRYELLEHKRWASSEEQRKYLKKRBEKKLKKKAKERKEKEMLEKLEKQK 578

Qy 550 --EDMECDDLKELPEPTPVKT--RLPPEIFGDALMVLFLNAFGEFLDQDEFDGVTTLE 605
Db 579 RYDEQELTG-KNLPAFLVDTPEGLPNTLFGDVAMVVEFLSCYSGLLLPDAQYP--ITAV 635

Qy 606 VLEALVGNDSG--PLCELLFPFLTAIFOAIAEBEEVAKEQITDADTKCSLSKSLDLD 663
Db 636 SLMEAL-SADKGGFLYLNRLVILLQTL--ODEIAE---DYGELGMKLSIPLT 688

Qy 664 SCTLSEILRLHILASGADVTSANAKYRYQKGGFDATDDACMELRLSNPSLVKLSSTSV 723
Db 689 LHSVSELVRLCLRRSDVOEESGSDTD-----DNKDSAAFEDNEVQDEFLEKLETSF 741

Qy 724 YDLTPGEKMKILHALCGKLLTLVSTRDFIEDYVDILRQAKQEFRELKAEQHRKEEAAA 783
Db 742 FELTSEKLOILTALCHRLMTYSVD-----HMETRQMSA 778

Qy 784 RIRKKEEKLKEQKQKMEKQEKLEDEQRNSTADISIGEEEREDFDTISBSKTEQK-- 841
Db 779 ELWKERLAVLKEENDKKRAEKQKREMEAKN-----KENGKVENGLKTDKRRKIV 829

Qy 842 ELDDQMTEDDDPGSHKGR---RGRKGONGKEFTRQIQINCVTRELITADAEERALK 897
Db 830 KFEQVDTAEADMTISAVKSRRLAIOAK-----KEREIQB-----REMKVKLERQA-B 876

Qy 898 QEHORKEKELLEK-IQSAIACNIF---PLGRDRMYRRYWF-PSIPGLFIEDYSGLT 951
Db 877 EERIRKHAAEKAFQEGIAKALVMERTPIGTDRNHNRYLWLFSDVPGLEIK---GWV 933

Qy 952 EDMLLPRPSSFQNNVQSDPOVSTKTGEP-----LMSESTSNIDQG-PRDHSVOL 1000
Db 934 HDSI---DYRFNHHCKDHTVS---GDDEYCPRSKXANLGNKASMTQHGTA TEVAVET 985

Qy 1001 PKPVHKPNRWCFYSSCQLDOLIEALNSRGRESALKETILQEKSRICAQLARSE---EK 1058
Db 986 TTPKQGNLWFLCDSQKELDELLNCLHPQGIRESQLKERL--EK-----RYQDIHS 1035

Qy 1059 PHSKDPQPSKPTYSRGRSSNAYDPSQMAEKQLELRDLFLDIEDRIYQGTIGAIV 1118
Db 1036 IHLARKPNLGLK-----SCDGNQ-----ELLNFRSDDLIEVATRLQKGLGYVEE 1080

Qy 1119 TDRHIWRSALSGRYELLSEENKENGIKTVNEVMEIDEQTKVIVK--DRLLGIKTE 1176
Db 1081 T-----SEPEA---RVISLEKL-----KDFGECVIALQASVIKKFLQGFMAPKQK 1122

Qy 1177 TPSTVSTNASTPQSVSVVHYLAMALFQIEGIERRELFKAPLDASDSGRSVKTVLDRWE 1236
Db 1123 RRKLQSEDSAKTEEVDE-----EKQVVEEAKVAS-----ALEKWK 1158

Qy 1237 SLLSASLSOVFLHSLTLDRSVIWSKILNARCKIKRKGDAENNVLCDCDRGHHTYCV 1296
Db 1159 AIRAQOTFSRMHVLLGMLDACIKWMSAENARCKVCPKGGEDDKLILCDECNKAPHLFCL 1218

Qy 1297 RPKLKTYPEGDWCFCECRPKQRCRLFRQRPSPLESDDEVEDSMGGEDEVDGDEEGOS 1356
Db 1219 RPALEYVDPGEWQCPACOPATARRNSGRNRYTEESASEDSDESDSEDEE---EEEBEE 1274

Qy 1357 EEBEYEVQEDDDSQEBEEVSLPKRGPQVRL-PVKT-RGKLSFSFSRSGQOQBPGRVPS 1414
Db 1275 EEDYEV-----AGLRLPRKTIRGHKHSVIPPAARSGRRPGRPKPH 1314

Qy 1415 RSQOSTPKTTVSSKTSRLSRKINSAPPTETKSLRIASRSTRSHSHGHPLOADVFEVLLSPR 1474


```
Db 1315 STRSQPK-----APPVD-----DAEVDL----- 1334
Qy 1475 KRRGRKSANTPNSPFPNFRVIATKSEQSRSVNIASKLSLOBSESKRRCKRQSP 1534
Db 1335 -----VLQTKRSRRQSL----- 1347
Qy 1535 SPTVLRSSRGQGVHLSAFEQLVVELVRHDDSWPFLKVSQVDPDYDIKKPIAL 1594
Db 1348 -----ELQCEETLHKIVKVFYFSPFPREPVRDEADYDVITHPMDF 1390
Qy 1595 NIIREKVNKCEYKLASEFDDIDELMFSNCFEYNPRNT 1631
Db 1391 QTVQNKCSGYSRVSQVEFLTDMKQVFNABVYNCRGS 1427

RESULT 13
US-10-702-148-29
; Sequence 29, Application US/10702148
; Publication NO. US20040063145A1
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042002
; CURRENT APPLICATION NUMBER: US/10/702,148
; PRIOR FILING DATE: 2003-11-05
; PRIOR FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 29
; LENGTH: 1531
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-702-148-29
```

```
Query Match 10.8%; Score 930.5; DB 15; Length 1531;
Best Local Similarity 23.2%; Pred. No. 3.9e-43;
Matches 399; Conservative 263; Mismatches 526; Indels 529; Gaps 68;
```

```
Qy 152 PLLHRKPFVRQKPPADLRDPDERVY-CKVTNFIHRHYDDFFERTILCSLWVSCAVTGRP 210
Db 3 PLLGRKPF-----PLVNPLPGEPPFTIPIHTQEAFTREBEYEARLERYSERIWTCKSTGSS 58
Qy 211 GITYQEALESEKARQNL-QSP-----EPLIIPVLVLTSTHRSRLHEICDDIFAVVKDQRY 266
Db 59 QLTHKEAEVEEOVAELKEEPPAWYKLVLEMVH-----HNTASLEKLVDTAWLEIMTKY 114
Qy 267 FVEETVEVIRNNGARLQCTIVLPL-----PS-----HONGFANGHV 303
Db 115 AVGECDPEVGEKMKVKVVIKHPLEKVBDEATEKSDGACDSSDKENSQIAQDHQ 174
Qy 304 NSVDGETIISDSDSETQS-----CSFQNGKKKDAIDPLL-FKYKQVPTKKBL 351
Db 175 KK-----ETVVKEDGREGSINDRARRSPKLPSTSLKKGKRWAPPKFLPHKYDV-----KLQ 227
Qy 352 HESAI---VKATOISRRKHLFSRDKLFLKQHC-----EPQGVIK-----IKASSLSTY 399
Db 228 NEDKIISNPADSLIRTERPPNKEIVRYFIRNALRAGTENAPWVVEDELVKYKSLPS- 286
Qy 400 KIAEQDSYFFPDPTPTIFSPANRRG-----RPPKRIHISQBDN----- 440
Db 287 -----KFSDFLLDPKYNTLNPSTKTKVTGSPDRKPSK-----SKTDNSSLSPNPKLWC 338
Qy 441 -VANKQTALA-----SYRSKATKER----- 458
```

```
Db 339 HVHLKLSUGSPLKVYKNSKNSKSPSEEHLEEMWKMMSPNKLHTNFHIPPKGPPAKPKGKHS 398
Qy 459 DKLLK-----QEMKSLAFKAKLR----- 479
Db 399 DKPLKAKGRSGKILNGQKSTGNSKSPKGLKTPPTKMKQMTLLDVAKGTKQKTRAPRNSG 458
Qy 480 -----EKADALBAKKKEKEDKKQR-----BEL 502
Db 459 GTPRTSPKPHKHLPPAALHLIAYYKENDKREDKRSALSCVISTKTARLLSSSEDRARLPBEL 518
Qy 503 KKIYEEE-----RLKKSEKRLKVEREKEREKLEB---EKRYKVEVLKQWSKPR 549
Db 519 RSLVQKRYELLEHKRWASMESEQRKEYLKKKEELKKLEKAKERKEKEMLEKQK 578
Qy 550 --EDMECDLDELPEPTPKT--RLPPEIFGDALMVLEFLNAFGSLFQDQEPFGVTL 605
Db 579 RYEDQELTG-KNLPAFLVDTPEGLPNTLFGDVAMVVEFLSCYSGLLLPDAQYP--ITAV 635
Qy 606 VLEALVGNDSG--PLCELLFFLTAIFQAIABEEBEEVAKELQTDATKGCGLSLDL 663
Db 636 SLMEAL-SADKGGFLYLNRLVILLQTLQTL--QDEIAB---DYGELGMLSEIPLT 688
Qy 664 SCTLSEILRLHILASGADVTSANAKYRYQKGGPDATDDACMELRLSNPSLVKLSSTSV 723
Db 689 LHSVSELVRLCLRRSDVQESGSDTD-----DNKDSAAFEDNEVODEFLEKLETSF 741
Qy 724 YDLTPGERKMTILHALCGKLLTLVSTRDFTEDYVDILRQAKQEFRELKAKQHRKEBEAAA 783
Db 742 FELTSEKLOILTALCHRLMTYSVD-----HMETQQMSA 778
Qy 784 RIRKKEEKLKEQEQMKKEQEKLEDQORNSTADISIGEBEERDFDTSIESKQTEQK-- 841
Db 779 ELWKERLAVLKEENDKKAERQKQKMEAKN-----KENGKVENGLGKTRDKRIV 829
Qy 842 ELDDQMTFDEDDPGSHKGR-----RGKRGQNGKFEFTROQINCVTVELLTADSEALK 897
Db 830 KFEQVDTEAEDMISAVKSRRLALIAQK-----KEREIQE-----REMKVLERQK-E 876
Qy 898 QEHORKEKELLEK-IQSAIACNTIF---PLGRDMYRYWIF-PSIPLGFIEEDYSGLT 951
Db 877 EERIRKHAARAEAFQGIKAKLVMRPTIGTDNHNRYWLFSDVEFGLFIEK---GWV 933
Qy 952 EDMLLPSPSSFNQVQSDPQVSTKTGP-----LMSESTSNIDQG--PRDHSVOL 1000
Db 934 HDSI-----DYRFNHHCDDHTVS---GDEDYCPRSKANLGKSNMTOHGTADEVAVET 985
Qy 1001 PKPVHKPNRWCFYSSCEQLDQIIEALNSRGHRESALKETILQESKRIQAQLARFSE--EK 1058
Db 986 TTPKQGNLWFLCDSQKELDELNCLHPQIGRESQKLERL--EK-----RYQDIIHS 1035
Qy 1059 FHSDKPPQDPSKPTVSRGRSSNAYDPSQMAEKOLELRDLFLDIEDRIYQGTILGAIKV 1118
Db 1036 IHLARKPNLGLK-----SCDGN-----ELLNFLRSDLIEVATRLQKGLGYVEE 1080
Qy 1119 TDRHWRSALESRGYELLSEENKENGIIKTVNEDEVEEMEIDEQTKVIVK--DRLGKITE 1176
Db 1081 T-----SEFEA---RVISLEKL-----KDFGECVIALQASVIKKFLQGFWAPKQK 1122
Qy 1177 TPSTVSTNASTPQSVSSVHYLAMALFOI EOGIERFLKAPLDASDSGRSYKTVLDWRRE 1236
Db 1123 RRKLQSEDSAKTEEYDE-----EKQMVVEAKVAS-----ALEKWK 1158
Qy 1237 SLLSSASISQVFLHLSLTDRLSVIWSKSTILNARCKIRKKGDAENNVCLDGCGRGHTTVCV 1296
Db 1159 AIREAQTFSRVHVLGLMDACIKWMSAENARCKVCPKKGDDDKULCDECNKAFHLFCL 1218
Qy 1297 RPKLKTVPEGDFCEPCRPKQRCRLSPRPSLESDESDVDSMGDEDDVDGDEEGQS 1356
Db 1219 RPALYEPDGEWQCPACQATARRNSGRNYTEESASEDSEDESEDEE---EEDDEE 1274
Qy 1357 BEEBEYEVQDDEDDQSEEBEEVSLPKRGRQVRL-PVKT-RGKLSSFSRSGQOQEPGRVPS 1414
```

Db	1275	BEEDYEV-----AGLRPRKKTIRGKHSVIPPAAAGRRGPKKPH	1314	Qy	441	-VANKOTLA-----SYRSKATKER-----	458
Qy	1415	RSQOSTPKTYSKTRSLRKINSAPPTETKSLRIASRSTRHSGPLOADVFWELLSPRR	1474	Db	339	HVHLKXLSGSPKVKXSKNSKSPSEHLEEMXOMSPNKLHTNFHIPKKGPPAKPGKHS	398
Db	1315	STRSQPK-----APPYD-----DAEVDL-----	1334	Qy	459	DKLK-----QEMKSLAFKAKLKR-----	479
Qy	1475	KRRGRKSANNTSPNSPFPNFRVJATKSSQSRSVNIAKSLQSESKRCRKRQSP	1534	Db	399	DKPLKAGRSKGIILNGQKSTGNSKSPKGLKTPKTKMQMTLLDMAKTQKTRAPRNSG	458
Db	1335	-----VLQTKRSSRQSL-----	1347	Qy	480	-----EKADALEAKKKEKEDKKR-----EEL	502
Qy	1535	SPVTLGRSSRGQGVHLSAFQOLAVELVRHDSWPFLKLVSKIQVDPYDYDIKKPTAL	1594	Db	459	GTPRTSSPKPHLPAAHLIAYYKENKREDKRSALSCVISTARTLSSSDRLARPEEL	518
Db	1348	-----ELQCEBILHKIVKVRFSWPPEPVRDEABEDYDVITHPWF	1390	Qy	503	KKIIVEB-----RLKXKEERLKVVEREKERKLE--EKKYVEYLKQWSKPR	549
Qy	1595	NIIREKVNCKBYKLASEFIDDIELMFSCFYNPNT	1631	Db	519	RSLVQKRYELLEHKRWASMSSEORKEYLKKRBEKKKKKKEKAKEREMERLEKQK	578
Db	1391	QTVQNKCGSYRSVOEFLTMKVQFTNAEYVNCGRS	1427	Qy	550	--EDMECCDLKELPEPTPVKT--RLPPEIFGDALMWLEFLNAFGELFDLQDFPDGVTLE	605
RESULT 14							
US-09-839-479-27							
; Sequence 27, Application US/09839479							
; Publication No. US20020039779A1							
; GENERAL INFORMATION:							
; APPLICANT: Jones, Michael H.							
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR							
; CURRENT APPLICATION NUMBER: US/09/839,479							
; CURRENT FILING DATE: 2001-04-20							
; PRIOR APPLICATION NUMBER: US 09/418,710							
; PRIOR FILING DATE: 1999-10-15							
; PRIOR APPLICATION NUMBER: PCT/JP98/01783							
; PRIOR FILING DATE: 1998-04-17							
; PRIOR APPLICATION NUMBER: JP 9/310027							
; PRIOR FILING DATE: 1997-10-24							
; PRIOR APPLICATION NUMBER: JP 9/116570							
; PRIOR FILING DATE: 1997-04-18							
; NUMBER OF SEQ ID NOS: 72							
; SOFTWARE: FastSeq for Windows Version 4.0							
; SEQ ID NO 27							
; LENGTH: 1527							
; TYPE: PRT							
; ORGANISM: Homo sapiens							
US-09-839-479-27							
Query Match 10.8%; Score 929.5; DB 9; Length 1527;							
Best Local Similarity 23.1%; Pred. No. 4.4e-43;							
Matches 397; Conservative 264; Mismatches 523; Indels 533; Gaps 68;							
Qy	152	PLLRKPPVRQKPPADLRPDSEVY-CVTVNEIRHYDDFFERTILCNLSVSCAVTGRP	210	Qy	952	EDMLPRPSPFQNVQSDPQVSTKTEP-----LMSESTSNIDQG-PRDHSVOL	1000
Db	3	PLLRKPP-----PLVNPLFGBEPFTTIPHTQEAFTREEYEARLERYSERIWTCKSTGSS	58	Db	930	HDSI-----DYRFNHHCKDHTVS---GDEDYCPRSKKANLGKKNASMTQHGTADEVAVET	981
Qy	211	GLTYQEALESKKARQNL-QSFP---EPLIIPVLYLTSLRHSRLHEICDDIFAYVKDQY	266	Qy	1001	PKPVHKPNRCFYSSCCQLDQLIBALNSRGHRESALKETLLOEKSRIKCAQLARFSE--EK	1058
Db	59	QLTHKEAWEQEVAELLKEFPAYWEKLVLEMHV-----HNTASLEKLVDTAWLEIMTKY	114	Db	982	TPKQGNLWFLCDSQKELDELLNCLHPQGIRESQLAERL--EK-----RYQDIHS	1031
Qy	267	FVEETVEIRNNGARLQCTILEVLP-----PS-----HONGFANGHV	303	Qy	1059	FHFSQKPPQDSKPTYSGRSSNAYDPSQMCABKQLELRDLFLDIEDRIYQGTGLAIKV	1118
Db	115	AVGECDEPVGKEMLVKVIKIHPLEKVEDEATEKKSQDGCDSPSSDKENSSQIAQDHQ	174	Db	1032	IHLARKENLGLK-----SCDGNQ-----ELLNFRSDLIEVATRLKQGLGYVEE	1076
Qy	304	NSVGETIILSDSDSETQS-----CSFQNGKKDAIDPLL-FKYKYQPTKKEL	351	Qy	1119	TDRHWSALBSRGYELLSEENKENGIIKTVNEVWEEMDEQTKVIVK--DRLLGIKTE	1176
Db	175	KK-----ETVVKDEGREINDRARRSPKLTSLKGERKWAPPKFLPHKYDV----KLQ	227	Db	1077	T-----SEFEA---RVISLEKL-----KDFGCVIALQASVKKKFLQGFMAPKQK	1118
Qy	352	HESAI---VKATQISRRKHLFSRDKLFLKHOHC-----EPQEGVIK-----IKASSLSTY	399	Qy	1177	TPSTVTSNASTPQSVSVVHYLAMALFQIEGIERRFLKAPLDASDSGRSVKTVLDRWRE	1236
Db	228	NEDKIISNVPADSLIRTERPNKEIVRYFIHNLRACTGENAPWVVEDELVKKYSLPS-	286	Db	1119	RRKQSDSASKEVDE-----EKKWVEAKVAS-----ALEKWK	1154
Qy	400	KIASQDSFYFPDPPFTFIFSPANRRG-----RPPKRIHSQEDN	440	Qy	1237	SLLSASLSQVFLHLSTLDRSVIWSKSIILNARCKICRKGDAENMVLCDGCDRGHTYCV	1296
Db	287	-----KFSDFLLDPYKVTNLNFTSKRNTGSPDRKPSKK---SKTDNSSLSPLNPKLWC	338	Db	1155	AIREAQTFSRMHVLLGLMDACIKWMSAENARCKVCPKKGEDDKLILCDECNKAPHLPLCL	1214

Db 1271 EEDIEV-----AGLRPRKTRGKSHVTPPAARSRRPGKPH 1310
Qy 1415 RSQOSTPTVSSKTRSLRKINSAPPTETKSLRIASRSTRSHGFLQADVFVLLSPRR 1474
Db 1311 STRRSQPK-----APPVD-----DAEDEL----- 1330
Qy 1475 KRRGRKSANNTSPNSFPNFVATKSEQSRSNVIAKSLQBSKRRCKRQSPPEP 1534
Db 1331-----VLQKRSRRQSL----- 1343
Qy 1535 SPVTLGRSSRGQGVHLSAEPQLAVVELVRHDDSWPFLKVLKIQVDPYDIKKPIAL 1594
Db 1344-----ELQKCEILHKLIVKTRFSWPFREPTRDEAEDYDVIITHPMDF 1386
Qy 1595 NIIREKVNKCEYKLASEFIDDIELMFNSCFYNPRINT 1631
Db 1387 QTVQNKCSGYSRSVQEFILDMQVFTNAEVNCRGS 1423

RESULT 15

US-10-376-537-27
; Sequence 27, Application US/10376537
; Publication No. US20030224405A1
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042001
; CURRENT APPLICATION NUMBER: US/10/376,537
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US/09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 27
; LENGTH: 1527
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-376-537-27

Query Match 10.8%; Score 929.5; DB 15; Length 1527;

Best Local Similarity 23.1%; Pred. No. 4.4e-43;

Matches 397; Conservative 264; Mismatches 523; Indels 533; Gaps 68;

Qy 152 PLLHRKPFVRQKPPADLRDEBVFY-CKVTNEIFRHYDDFFERTILCNLSVMSCAVTGRP 210
Db 3 PLUGRKPP-----PLVNLPGEEPFTIPIHTQBAFTRIEYEARELERYSERIWTCKSTGSS 58
Qy 211 GITYQEALESEKKARQLN-QSEF---EPLIIPVLYLTSLTHRSRLHEICDDIFAYVKDQY 266
Db 59 QLTKEAEQEEQEAELLKEEFPANVEXLVLEWVH---HNTASLEKLVDAWLEIMTKY 114
Qy 267 FVEETVEVRNNGARLQCTILEVLP-----PS-----HONGFANGHV 303
Db 115 AVGECDFEVGKEKMLKVIKIHLEKVEBATEKKS DGACDSPSKENSSQIAQDHQ 174
Qy 304 NSVDGETIISDSDSETOS-----CSFQNGKKDAIDPLL-FKYKVQPTKKEL 351
Db 175 KK---ETVVKEDEGRRESINDRARRSPKLPSTLKKGERKWAPPKFLPHKYDV-----KLQ 227
Qy 352 HESAI---VKATQISRRKHLFRDKLKLFLKQHC-----EPQEGVIK---IKASSLSTY 399
Db 228 NEDKIISNVADSLRTERPPNKEIVYFIHNLRACTGENAPVWVEDELVKYSLPS- 286
Qy 400 KIAEQDFSYFFDPDPTTFISPANRRG-----RPPKRIHISQBDN----- 440
Db 287 -----KFSDFLLDPKYMTLNPSTKRKNTGSDPKPSKK---SKTDNSSLSPNKLWC 338

Qy 441 -VANKQTIA-----SYRSKATKER----- 458
Db 339 HVHLKSLSGSPLKVNKSNKSPBEHLEEMWOMSPNKLHTNPHIPKKGPPAKPGKHS 398
Qy 459 DKLLK-----QEMKSLAFEXAKLKR----- 479
Db 339 DKPLKAKGRSGKILNGQKSTGNSKSPKGLTKPTKMKMQMTLLDMAKGTQKQWTRAPRNG 458
Qy 480-----EKADALEAKKKEKEDKEKKR-----BBL 502
Db 459 GTPRTSSKPHKLPPAALHLIAYYKENDKREDKRSALSCVISKARTLRLSSDEARLPEEL 518
Qy 503 KKIIVEE-----RLKKBEKERLKVEREKERELRE--EKRKYVEYLKQWSPR 549
Db 519 RSLVQRYELLEHKRWASMEEQKEYLKKREBELKKLKEKAKERERKEMLERLEKQK 578
Qy 550--EOMECDDLKELEPPTPKT--RLPPEIFGDALMVLFLANAFGLFDLQDEFDPGVTL 605
Db 579 RYEDQELTG-KNLPAFLVDTPEGLFNTLFGDVAMVVEFLSCYSGLLLPDAQYP--ITAV 635
Qy 606 VLEBALVGNDSGELPCCELLFPFLTAIFAIAE--EEBVAKEQLTDAOTKGCSLKSLDLD 663
Db 636 SLMEAL-SADKGG-----FLYLRVVLVILQTLQDEIAE-----DYELGNKLSEIPLT 684
Qy 664 SCTLSEILRLHILASGADVTSANAKYQKRGGFDDACMELSLSPSLVKLSSTVS 723
Db 685 LHSVSELVRLCLRRSDVQEESEGSTD-----DNKDSAAFEDNEVDQEFLEKLETSF 737
Qy 724 YDLTFGEKQKILHALCGKLLTLVSTRDFIYDVILLROAKQEFRELKABQHRKEBEAAA 783
Db 738 FELTSEKLOILTALCHRLMTYSVQD-----HMETRQOMSA 774
Qy 784 RIRKKEBKLKEQEKQKKEKEDKEDORNSTADISIGEREERDFDTSIESKDETEK-- 841
Db 775 ELWKERLAVLBENDKGRKQKQKEMPAKN-----KENGKVENGLGKTDKRRIV 825
Qy 842 ELDDQMFTEDEDDPGSHKRRGR---RGKRGQNGKFEFTRQEQINCVTRRELITADSEELK 897
Db 826 KFEQVDTEAEDMISAVKSRLLAIQAK-----KEREIQE-----REMVKLRSQA-E 872
Qy 898 QEHORKEKELLEK-IQSAIACNIP-----PLGRDRMYRYWIF-PSIPLGLFIEEYSGLT 951
Db 873 BEIRKXKAAAEKAFQEGIAKAKLVMRRTPICTDRNHNRYWLFSDVDEVPGLFIEK---GWV 929
Qy 952 EDMLLPRESSFQNNVQSDPOVSTKTGP-----LMSESTSNIDQG-PRDHSVOL 1000
Db 930 HDSI-----DYRFNHHCKDHTVS---GDEDYCPRSKKANLGNASMTQHGTAETAVET 981
Qy 1001 PKPVHKPNRWCYFSCQOLDQLEALNSRGHREGALKETTLQEKSRICAQLARFSE--EK 1058
Db 982 TTPKQGNLWFLCDSQKELDELNCLHPQGIRESOLKERL--EK-----RYODIHS 1031
Qy 1059 FHFSDKPOPDSPKPTVSRGRSNAYDPQSCMAEQLELELRDLFLDIEDIRIVQGTGAIKV 1118
Db 1032 IHLARKPNLGLK-----SCDGNQ-----ELNFLRSDLIEVATRLQKGLGYVEE 1076
Qy 1119 TDRHWRSALESRGYELLSEENKENGIIKTVNEDEVEEIDEQTKVIK--DRLIGIKTE 1176
Db 1077 T-----SEFEA---RVISLEKL-----KDFGECVIALQASVIKFLQGFMAPOK 1118
Qy 1177 TPSTVSTWASTPOSVSSVVHYLAMALFOIEQIERRFLKAPLDASDSGRSYKTVLDRRE 1236
Db 1119 RRKQESDASATEEVE-----EKKQVEEAKVAS-----ALEKWT 1154
Qy 1237 SLLSASLSQVFLHLSTLDRSVIWSKSLNARCKICRKGDAENMVLCDGCDRGHTTVCV 1296
Db 1155 AIREAQTFSRMHVLLGMLDACIKWDSAEANARCKVCPKKGDDKLLCDECNKAPHLFCL 1214
Qy 1297 RPKLKTVPGEQWDFCEPKPCORCLRLSPRQPSLESDEDDVSMGEGEDVDGDEEQQS 1356
Db 1215 RPALYEVDPGEWQCPACOPATARRNSRGRNTYTESASDESDEDEDEE-----BEEEEE 1270

```

Qy 1357 EEBEYVEQEDDDSQEEBEVSLPKRGQPVTL-PVKT-RGKLSSESSRRQQQEPGRYPS 1414
Db 1271 EEDYEYV-----AGLRLPRKTIIRGSHSVIPPAARSGRRPGKKPH 1310
Qy 1415 RSQOSTPKTVSSKTRSLRKINSAPPTETKSLRIASRSTRHSHGPLQADVFVVELLSPPR 1474
Db 1311 STRKSQPK-----APPYD-----DAEVDL----- 1330
Qy 1475 KRGKRSANNTPENSPNPFNFRVIATKSSBQSRSVNIATSKLSQESKRCRKROSPPEP 1534
Db 1331 -----VLQTKRSSRRQSL----- 1343
Qy 1535 SPVTLGRSSRGQGVHLSAFEOQLVVELVRHDDSWPFLKLVSKIQVDPYDYDIKKPIAL 1594
Db 1344 -----ELQKCEILHKIVKIRFSWPFREBVRDEADYDVIITHPMDF 1386
Qy 1595 NIIREKVNKEYKLASEFFIDDIELMFNSCFEYNPRNT 1631
Db 1387 QTVQNKCCSRSYRQVEFLTDMKQVFTNAEYVNCGRS 1423

RESULT 16
US-10-702-148-27
; Sequence 27, Application US/10702148
; Publication No. US20040063145A1
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042002
; CURRENT APPLICATION NUMBER: US/10/702,148
; CURRENT FILING DATE: 2003-11-05
; PRIOR APPLICATION NUMBER: PRIOR APPLICATION NUMBER: US/09/839,479
; PRIOR FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 27
; LENGTH: 1527
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-702-148-27

```

Qy	400	KIAQOPSPYFPDDPTTFIPSPANRRG-----RPPKRIHISQEDN-----	440
Db	287	-----XFSDFLLDPYKYMTLNPCKTKNTGSPDRKPSKK-----SKTDNSILSSPLNPKLWC	338
Qy	441	-VANKOTLA-----SYRSKATKR-----	458
Db	339	HVHLKXSLSGSPLKVKXNSKNXSKSPSEHLEBMMQOMSPNKLHTNFHIPKKGPPAKPKGKHS	398
Qy	459	DKLLK-----QEBMSJLAPEKAKLKR-----	479
Db	399	DKPLKAGRSKGILNGOKSTGNSKSPKGLKTPKTKQWTLDDMAKGTQKMTAPRNSG	458
Qy	480	-----EKADALEAKKKEKEDKPKKR-----	502
Db	459	GTPTRSKPHKLPPAALHLIAYKENKOREDKRSALS CVISKARTALLSSEDRARLPEEL	518
Qy	503	KKIVEBE-----RLKXKEERLKVBEREKREKLRE-----EKRKVVEYLKQWSKPR	549
Db	519	RSLVOKRYELLEHKRWASMSSEQRKEYIKKKEELKCKLKEKAKERREKEMLERLEKQK	578
Qy	550	--EDMECDDLKELPETPVKT--RLPPEIFGDALMWLEFLNAPGELFDLQDFPDQGVTL	605
Db	579	RYEDELGT-KNLPARFLVDTPGELPNTLFGDVAMVVEFLSCVSGLLLPDAQVP--ITAV	635
Qy	606	VLEEALVGNDSGELCELUFFFLTAIFAIAEB--EEEVAEQLTADTKGCSLSKSLDLD	663
Db	636	SLMEAL-SADKGG-----FLYLNRLVILLQTLQDEIAE-----DYGELGMKLSBIPLT	684
Qy	664	CTLSEILRLHILASADVTSANAKRYQKRGFGDATDDACHMELRLSNPSLVKLLSSTSV	723
Db	685	LHVSLELRLCLRRSDVQBESEGSTD-----DNKDSAAFEDNEVQDFLEKLETSBF	737
Qy	724	YDLTPGOMKILHALCGKLLTLVSTDFIEDYVDILRQAKQBFRELKABOHRKEREAAA	783
Db	738	FELTSSEKQLTALCHRIIMTVSD-----HMETROOMSA	774
Qy	784	RIRKREKLEKEQORMKQEKLEKEDQORNSTADISIGEEREDPDTSESOTEQK--	841
Db	775	ELWKERLAVKEENDKKRAEKQKRKEWAKN-----KENGKVENG LGTKTRDKRIV	825
Qy	842	ELQDMFTEDDDPGHKHGR-----RGKGQNGFKFTRQEQINCVTRELLTADAEALK	897
Db	826	KFEPQVDTEADNIMSAVKRRRLAIAQK-----KEREQE-----REMKVKLERQA-E	872
Qy	898	QEHORKEKLEK-IQSACTACTWIF---PLGRDRMYRYWIF-PSIGPLFTIEEDYSGLT	951
Db	873	EEIRKXKAAEAKAFEGIAKAKLVMRPTPIGTDRHNHNYWLFSDVPGFLTEK---GWV	929
Qy	952	EDMLPRPSSFQNNVOSQDPQVSTKTGP-----LMSESTSNIDQG--PRDSVOL	1000
Db	930	HDGI-----DYRFNHCKDHTVS--GDIEDYCPRSKKANLGKNASMTQHGTADEVAVET	981
Qy	1001	PKPVHPKPNRCFYSSCEQLDOLIEALNSRGHRESALKETILQEKSRICAQLARFSE--EK	1058
Db	982	TTPKQGNLWFLCDSQKDELNCLHLPQIGRESQKERL--EK-----RYQDIIHS	1031
Qy	1059	FHPSDPQPPSKFTYSGRRSSNAYDPSQCAEQELRLRLDFLLDIEDRIYQCTLCAIKV	1118
Db	1032	IHLARPKNLGLK-----SCDGNQ-----ELNLFRLSDLIEVATLQKGLGVVEE	1076
Qy	1119	TDRHIWRSALSGRYELLSEENKENGIIKTNVEDVBEIMEIDQTKVIVK--DRLLGIKTE	1176
Db	1077	T-----SEFEA--RVISLEKL-----KDFGCVIALQASVIKKFLOGFMAPKQK	1118
Qy	1177	TPSTVTNASTPOSVSSVHYLAMALFQIEQIERFLKAPLDASDSGRSYKTVLDRWE	1236
Db	1119	RRXLQSDSAKTEVD-----EKQWVEAKVAS-----ALEKWK	1154
Qy	1237	SLSSASLSQVFLHLSLDRSVIWSKSIILNARCKICRKKGDAENMWLVCDGRGHHTYCV	1296
Db	1155	AIREAOTFSMMHVLGMLDACIKWDWSAENARCVCKPKGEDDKLLTLCDBCNKAFHLFCL	1214
Qy	1297	RPKLKTVPEGDWFCPCPRKQRCRRLSFRQPSLESDVEDVDSMGGEDEVGDDEEGOS	1356

Qy	1301	KTVPEGDWFCPECPKQCRRLLSPQRPSLESDEVEDSMGGEDEVGDGEEEGSEEEE	1360
Dd	1218	YEVDPGEWQPACQAPATARNRSGRNTEBSASED-----ESDE-EEEEEEEEEEED	1272
Qy	1361	YEVEODEDDQEEBEVSLPKEGRPOVFL-PVKT-RGKLSSFSRSGQQOEPGRYPERSQ	1418
Dd	1273	YEV-----AGLRLPRKTLRGHSHVIPPAARSGREPGKKPHSTR	1312
Qy	1419	STPKTTVSKTGSLRKINGSAPPETKSLRIASRSTRHSHPGLQADVVELLSPPRRKRGG	1478
Dd	1313	SQPK-----APPVDAE-----	1323
Qy	1479	RKSANNTPENSPNPNFRVIATKSSSEQSRNSIATKLSLOESKRRCKRQSPSPVT	1538
Dd	1324	-----VDELVLQT-----	1331
Qy	1539	LGRSSRGQGVHELSAFEQLWVELVRHDSWPFLKLVSKIQVPDYDIITKKPIALNIIR	1598
Dd	1332	--KXSSRRQS--LELOKCEBILHKIKVYRFSPWFREPVTDEAAEDYYDVITHPMDFTQVQ	1387
Qy	1599	EKNVCYKULASEFIDDIELMFSCFNPNPT	1631
Dd	1388	NKCSGYRSVQOEPLTKMGOVFTHAENVNCRGS	1420

```

RESULT 19
US-10-702-148-68
; Sequence 68, Application US/10702148
; Publication No. US20040063145A1
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042002
; CURRENT APPLICATION NUMBER: US/10/702,148
; CURRENT FILING DATE: 2003-11-05
; PRIOR APPLICATION NUMBER: PRIOR APPLICATION NUMBER: US/09/839,479
; PRIOR FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 68
; LENGTH: 1525
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-702-148-68

```

Query Match	10.7%;	Score 926;	DB 15;	Length 1525;
Best Local Similarity	23.2%;	Pred. No. 6.9e+43;		
Matches 397;	Conservative 261;	Mismatches 527;	Indels 528;	Gaps 67;
Qy 152	PLLHRKFFVRQKPADLRPDEEVY-CKVTNEIRHHYDDRPFTILGNLSLWSCAVTGRP	210		
Db 3	PLLRKPF-PLVNLPGEPFTTPTQEARTRREYEARLERSERLWTCKSTGSS	58		
Qy 211	GLTYQEALESKKARQNL-QSPF---BPLIIPVLYLTSLTHRSRLHEICDDIFAYVKDXY	266		
Db 59	QLTHKEAWEESEQVAELLKEEFPAWYKLVLEMVH---HNTASLEKLVDTAWLMEITKY	114		
Qy 267	FVRETVSVIRNGARLOCTILEVLP-----PS-----HONGFANGHV	303		
Db 115	AVGECEDFVGEKRMVKVVIHPLKVDDEATEKKSDBGACDSFSSDKENSSQIAQDHQ	174		
Qy 304	NSVDGETIIISDSDDSETQS-----CSFQNGKKKDAIDPLI-FKYKVQPTKKEL	351		
Db 175	KK---ETWVKDEGRGESSINDRRSPKPLTSLKKGRKWAPKPFPHKYDVF---KLQ	227		

Qy	352	HESAI---	VKATOISRRKHLSRDKLKLFLKQHC-----EPOEGVIK---IKASLSSTY	399
Db	228	NEDIISNVPADSLIRTERPPNKEIIVRYPIFRHNALRAGTENAPWVVEBELVKYSILPS-	286	
Qy	400	KIAEQDPSFPDDPTTFPSPANRRG-----RPKRIHIISOEDN-----	440	
Db	287	----XPSDFLLDPYKYMTLPSTPKKNTGSPDRKPSKK---SKTDNSSLSPLNPCLMC	338	
Qy	441	-VANOKTLA-----SYRSKATKER-----	458	
Db	339	HVHLKKSLSGPLKVKNKSNKSPSEHLSBMWKMSPNKLHTNPHIPKKGPPAKKPCKHS	398	
Qy	459	DKLLK-----OEEMKSLAFPAKLKR-----	479	
Db	399	DKPLKAKGRSGILNGOKSTGNSKSPKGLTKPTKWQMTLLDMAKTQOMTRAPNSG	458	
Qy	480	-----EKADALEAKKEKEDKEKR-----	EEL 502	
Db	459	GTPRTSGPKHHLPAAALHLIAYKENKOREDKRSAISCVI SKTABLLSSEDRLAPEEL	518	
Qy	503	KKITVEE-----RLKKEKEERUKVEREKEREKLE--EKRKYVEYLKOWSKPR	549	
Db	519	RSUVQRXYELLEHKKWASHMSEBQRKEYLUKKREELKCKLKAVAKERREKEMLERLEKQK	578	
Qy	550	--EDMECDOLKELPEPTPVKT--RLPPEIFGDALMVLEFNAFGELFDLODEPPDGVTLE	605	
Db	579	RYEDQELTG-KNLPAFRLVDTPBGLNLTLPFGDVAMVVFSLSCVSGLLPDAQYP--ITAV	635	
Qy	606	VLBEALVGNDSEGPLCELLFFFTAITPAIAEB--EEVAKBOLTADTDYGCSLKSLLDLDD	663	
Db	636	SLMEALA--DKGG-----FLYNLRVLVILLQTLLQDEIAB---DYGLGMKMLSIPLT	683	
Qy	664	SCTLSETLRLHILASGADVTSANAKYRQKRGGFDTDDACMELRLSNFSLVKCLKSSTS	723	
Db	684	LHVSSELVRCLRRSDVOQSEGSDDT-----DNKDSAAFEDNEVQEDEFLEKJTSFE	736	
Qy	724	YDITPGCKMKILHALCGKLLTVLSTRDFIEDYVDILRQAQEPRELKAEQHRKEREEAAA	783	
Db	737	FELTSSEKIQLTALCHRILMTYSVQD-----HMETROQMSA	773	
Qy	784	RIBRKBEELKEQEQKQKQKQKQKQKEDDEBORNSTADISIGEBBERDDPTSIESKDTSQK--	841	
Db	774	ELWKERLAVLKEENDKKAEEKQKQKKEWEAKN-----KENGKVENGLGKTDRKKRIV	824	
Qy	842	ELDQDAFTDEDDEDPGSHKGR-----RGKRGQGFKFTREQINCINCVTRELLTADBEALK	897	
Db	825	KFPQVDVTTAEDMISAVKRRLLIAIOAK-----KEREIQE-----REMVKVLEKQA-E	871	
Qy	898	QEHORKEKELEK-IQSAIACTNIF----PLGRDRMYRWYWP-PSIPGLFIEEDYSGLT	951	
Db	872	EERI RKHKA AEAKAFOGI AKAKLVNRRPIGTDRNHNYWLFSDEVPGFLIEK---GWV	928	
Qy	952	EDMLPRPSPFNNOVQSPQVVS-----TKTGEPIMSESTSNIDQG-PRDHVSVOLBK	1002	
Db	929	HDSI-----DYRFNNHHCKDHTVSGCEDYCPRSKKAN-LGKNASMNTHOGTATEVAVETT	982	
Qy	1003	PVHKPNRCWFYSCEOOLDOLI EALNSRGHRESNAKETLLOEKSRICAQLARSEEKFHES	1062	
Db	983	PKGOQLNWFLCDSQKSELDLNCCLHPQOGRESQLKERLEKYODIIHSL-----HLA	1034	
Qy	1063	DKQPQDSKPTYSRGRSSNAVDPQMCAEQLELRDLFDLEDRIYQGTGLGAIKVTDNRH	1122	
Db	1035	RKNLGLK-----SCDGNQ-----ELLNFLASDLIEVATRLKQKGLGVET---	1076	
Qy	1123	IWRSALESGRYELLSENKENGIIKTWNEDVBEMEIDEQTKVIVK--DRLLGIKTFPST	1180	
Db	1077	---SEFEA---RVISLEKL-----XDFGEVIALQASVIKKPLOGFMAPKQKRKL	1121	
Qy	1181	VSTNASTPOSVSVVHYLAMALFOIBQGIERRFLKAPLDASDSGRSKYTVLBWRRESLIS	1240	
Db	1122	QSDSDAKTEEVDE-----EKKMVEBAKVAS-----ALEKMTAIRB	1157	

Qy	1397	SSSFSSRGQQQEPGRYPSRSQQSPKTTVSSKTSRSLRKINSAPPTETKSLRIASRSTRH	1456
Db	1823	-----SASTSSSL-----	1830
Qy	1457	SHGFLQADVVELLSPRKBRGRKSANNTPENSFPNFRVIATKSEQSRSVNIASKLS	1516
Db	1831	-----KRGKND-----LQCRKMEENTSIN-----LS	1851
Qy	1517	LQESSEKRRCKRKQSPSPSVTLGRRSSGRGGVHLSAFEQLVVELVHRHDDSWPFLKLV	1576
Db	1852	QKESFTSVKPKPRDSSK-----DLALCSMLTETMETHEDAWPFLPV	1993
Qy	1577	SKIQVPPYDIIKKPIALNIREKVNKECYKLASEFTDDIELMFSNCFEYNPNTSEAKA	1636
Db	1894	NLKLVPGYKKVKKPMDFTSIREKLSSGGYPNLETFALDVRLVFDNCETFNEDSDIGRA	1953
Qy	1637	GTRLQAPFF 1644	
Db	1954	GHNWKYF 1961	
RESULT 21			
US-10-376-537-21			
; Sequence 21, Application US/10376537			
; Publication No. US20030224405A1			
; GENERAL INFORMATION:			
; APPLICANT: Jones, Michael H.			
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR			
; FILE REFERENCE: 06501-042001			
; CURRENT APPLICATION NUMBER: US/10/376,537			
; CURRENT FILING DATE: 2003-02-28			
; PRIOR APPLICATION NUMBER: US/09/418,710			
; PRIOR FILING DATE: 1999-10-15			
; PRIOR APPLICATION NUMBER: PCT/JP98/01783			
; PRIOR FILING DATE: 1998-04-17			
; PRIOR APPLICATION NUMBER: JP 9/310027			
; PRIOR FILING DATE: 1997-10-24			
; PRIOR APPLICATION NUMBER: JP 9/116570			
; PRIOR FILING DATE: 1997-04-18			
; NUMBER OF SEQ ID NOS: 73			
; SOFTWARE: FastSeq for Windows Version 4.0			
; SEQ ID NO 21			
; LENGTH: 1972			
; TYPE: PRT			
; ORGANISM: Homo sapiens			
US-10-376-537-21			

Query Match

[illegible][illegible]

```

Db 1831 -----KRGKND-----LQKXKMBEINTSIN-----LS 1851
Qy 1517 LOESKRCRCRQSPSPVTLGRRSSRGQGVHLSAFQLVVLVVRHDDSWPFLKLV 1576
Db 1852 QKESFTSVKPKRDDSK-----DLALCSMILTEMETHEDAWPFLPV 1893
Qy 1577 SKIOVPDYDIKIPALNIIREKVNKCEYKLASEFIDDIEMFNSCFEYNPRTSEAKA 1636
Db 1894 NLKLVPGYKVIKPMDFSTIREKLSQCPNLETALDVLVDFDNCETFNEDSDIGRA 1953
Qy 1637 QTRLQAFF 1644
Db 1954 GHNMRYKF 1961

RESULT 22
US-10-702-148-21
; Sequence 21, Application US/10702148
; Publication No. US20040063145A1
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042002
; CURRENT APPLICATION NUMBER: US/10/702,148
; CURRENT FILING DATE: 2003-11-05
; PRIOR APPLICATION NUMBER: PRIOR APPLICATION NUMBER: US/09/839,479
; PRIOR FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US/09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 21
; LENGTH: 1972
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-702-148-21

```

```

Query Match      8.4%; Score 726.5; DB 15; Length 1972;
Best Local Similarity 21.4%; Pred. No. 1.3e-31;
Matches 335; Conservative 195; Mismatches 425; Indels 613; Gaps 53;

Qy 420 SPANRRRGRRP-----KRIHQSDNVA--- 442
Db 664 SRMRKGRPNVGNNAEFLDNADAKLLRKLQAQBIARQAQIKLLRKLQKQEQARVAKA 723
Qy 443 --NKQTLASYRSKATKERRDKLLKQEE-----MKSIAFE-----KAKLKREKADA- 484
Db 724 KQQAIAAAEKKRQKEQIKIMKQOEKIKRIQOIRMEKELRAQOILKAKKKKEAANAK 783
Qy 485 -LEAKK--KEKE-----DKEKKRELKXIVVEERLKKKEERLKVREKE-- 527
Db 784 LLEAKRTKEKEMRRQAVLLKHQOERRRRQHMMMLMKAMEARKKAAEKERLQKRDKEKR 843
Qy 528 --REKLREKKRYEYLKQWSKPRDECDLKEPPE--PTPVKTRLPPEIFGDALMVLEF 584
Db 844 LNKKERKLEORLEEMAKELKPNEDCLADQKPELPRIPLGVLSGSTFSDCLMVQVF 903
Qy 585 LNAFGEL--FDQDEFPDQVTVLEVEALVG--NDSEGPLCELLFFFLTAIFAIAEBEEE 641
Db 904 LRNFQKVLGFDVNDVNP-----LSVLQEGLLNIGDSMGVQDILLVRLLSA---AVCDPGLI 957
Qy 642 VAKQELTDADTKGCSLSKSLDLDSDCTLSILRLHLILASGADVTSANAKRYQKRGFDATD 701
Db 958 TGYKAKT---ALGEBHLNVGNVRDNVSEILQIF----- 987
Qy 702 DACMELRLSNPLVKKLSSTSVYDLTPQEKMKILHALCGKLL---TLVSTRDFIEDVYDI 758

```

```

Db 988 --MEAHGOGOTELTESLTKAFQAHTPAQKASVLAFLINELACSKSVVSEIDKNIDYMSN 1044
Qy 759 LRQAKQBFRELKABQHRKEREAAAARTRKKEEKLKEQEQKMKKQKELKDEDEBORNSTAD 818
Db 1045 LRRDKWV-----EGKLR-----KLRIIHAKTGRDTSGG 1075
Qy 819 ISIGEE-----REDFDTSESQTEKELDQDMFTEDDEDDPGSHKGRGK 865
Db 1076 IDLGEHQPLGTPPGRKRRKGGSDYDDDDDDSDQDDEDEDEDEDEDEDEDEDEDEDEDEDE 1131
Qy 866 RGQNGPFKEFRQEQINCVTRELLTADBE-----ALKQBHQKKEKELLEKIQS-----A 914
Db 1132 KTD-----ICEDEDEGQAASVEELEKQIEKLSKQSQYRRKLPDA 1172
Qy 915 IACTNIFPLGRDRMYRWIWPSPGLFIE--EYSGLTEDMLLPRSSFNQNVQSQDPQ 972
Db 1173 SHLSRVNFGPDYRRRYWILPRGGIFVEGMESEGLEETAKEREKAKAESVQIKBEEM 1232
Qy 973 VSTKTGRLMSESTSNIDQG-----PRDHSVQLP 1001
Db 1233 PET--SGDSLNCSTDHCEQKEDLKEKONTNLFLOKPGSFKLSKLELVAKMPPSEVWTP 1291
Qy 1002 KP-----VHK-----PNRWCYSSCEQ 1018
Db 1292 KPNAGANGCTLSYQNSGKHSILGVSQSTATQSNVEKADSNLNTGSSGPGK--FVSPLPN 1349
Qy 1019 LDQIIEALNSRGHRESAL-----KETLLQEKSRICQALARPSEKHFHSDKPOPD----- 1068
Db 1350 -DQLLKTLEKNRQWFSLLPRTPCDDTSLTHADMSTASLVTPOQPPSPSPAPLGS 1408
Qy 1069 -----SKPTYSRG-----RS 1078
Db 1409 SAQNPVGINFPALSPLOVKGVSMMGLQFCGWPTGVVTSNIPFTLSVPSLGSGLSGN 1468
Qy 1079 SNAYDPSQMAEQK-----LEL-RLRDF-----LMDI 1104
Db 1469 GNSFLTNSVASSKSESPVPQNEKATSAQPAAVEVAKPVDPPSPKPIPEEMQFGWRIIDP 1528
Qy 1105 EDRIYQGTLGAI-----KVTDH---TWRSALSGRYELL-S-ENKENG 1145
Db 1529 ED-----LKALLKVLHRLGIREKALQKIQKHLDYITQACLKNKQVAIIELENEENQV 1582
Qy 1146 IKTV--NEDVEEMIEDQTKVIVKDRLLGKTEPSTVSTNASTPOSV----- 1192
Db 1583 TRDIVENWSVEEQAMWDLVLQQVEDLERRVAGASLOVKGWMCPEFASEREDLVYFEHK 1642
Qy 1193 -----SVVHYL-----AMALFQIEQGIERRFLK--APLDAS 1221
Db 1643 SFTKLCKEHDGEFTGEDSSAHALERKSDNPLDIAVTRLADLERNIERRIEEDIAP--- 1698
Qy 1222 DSGRSYKTVLDRWRRESLLSSASLSQVFLHLSTLDRSVIWSKSLNARCKICRKKGDAENM 1281
Db 1699 -----GLRVWRALSEARSAAQVALCITQLOKSIWEKSIKMKVYCOICRKGDNEELL 1750
Qy 1282 VLCDGCDRGHHTYCVRPKLKTVPEGDFCPECRPKQ-----RCRRLSFROPSPSLEDV 1336
Db 1751 LLCGCDKGCHTYCHRPKTIITIPDGFPCACIAKASGQTLKIKKLVHVKKHTNESK-- 1808
Qy 1337 EDMSGGDEDDVDGDEEBEGQSEEBEYVEQDEDDSQEBEEVSLPKRGRPQVRLPVKTRGKL 1396
Db 1809 -----GKKVTLTGDE-----DED----- 1822
Qy 1397 SSSPSSRGQOQEPGRYPSPRSQQSTPTKTVASKTGRSLRKINSAPPTETKSURIASRSTRH 1456
Db 1823 -----SASTSSSL----- 1830
Qy 1457 SHGLQADVVELLSPRKBRGRKSANNTPENSPNFNFRVIAKTSQESRSVNIASKLS 1516
Db 1831 -----KRGKND-----LQKXKMBEINTSIN-----LS 1851
Qy 1517 LOESKRCRCRQSPSPVTLGRRSSRGQGVHLSAFQLVVLVVRHDDSWPFLKLV 1576
Db 1852 QKESFTSVKPKRDDSK-----DLALCSMILTEMETHEDAWPFLPV 1893

```


Db 1950 AGHNMRYF 1958

RESULT 24

US-10-376-537-72

Sequence 72, Application US/10376537

Publication No. US20030224405A1

GENERAL INFORMATION:

APPLICANT: Jones, Michael H.

TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR

FILE REFERENCE: 06501-042001

CURRENT APPLICATION NUMBER: US/10/376.537

CURRENT FILING DATE: 2003-02-28

PRIOR APPLICATION NUMBER: US/09/418.710

PRIOR FILING DATE: 1999-10-15

PRIOR APPLICATION NUMBER: PCT/JP98/01783

PRIOR FILING DATE: 1998-04-17

PRIOR APPLICATION NUMBER: JP 9/310027

PRIOR FILING DATE: 1997-10-24

PRIOR APPLICATION NUMBER: JP 9/116570

PRIOR FILING DATE: 1997-04-18

NUMBER OF SEQ ID NOS: 73

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 72

LENGTH: 1969

TYPE: PRT

ORGANISM: Homo sapiens

US-10-376-537-72

Query Match 8.2%, Score 704, DB 15, Length 1969;

Best Local Similarity 21.1%, Pred. No. 2.4e-30;

Matches 331; Conservative 196; Mismatches 426; Indels 616; Gaps 53;

Qy 420 SPANRRGRPP-----KRTHSQEDNVA----- 442

Db 662 SRMRKGRPPNVGNABFLDNADAKLRKLOAQBARQAAQIKLRKLQKQBARVAKEAK 721

Qy 443 -NKQTASYRSKATKRDKLLKQEE---MKSFAPE-----KAKLKREKADA-- 484

Db 722 KQAIMAAEKRKQKQEKIKHQEKIKRIQIIRMEKELRAQQILEAKKKKKEAANAAL 781

Qy 485 LEAKKKEKE-----DKEKKREBKIKVBEERLKKKEERKAKVEREKE--- 527

Db 782 LEAEKRIKERWRQAVLLKQERERRRQHMMLKWAERKKAERLQKQEKDEKRL 841

Qy 528 -REKLREKRYEVLKQWSXPREDMECDLKELP-PTPVKTRLPPEIFGDALMVLEFL 585

Db 842 NKERKLEQRLEMAKELKPNEDMCIADQKPLPELPRIPLGLVLSGSTFSDCLMVVQFL 901

Qy 586 NAFGEL--FDLQDFPDGVTLEVEALVGN--DSEGPLCELLFFFLTAIFAABESEE 641

Db 902 RNFGKVLGFDVNIQVFN---LSVLOEGILLNIGDSMGVQDILLVLLSA---AVCDPLGI 955

Qy 642 VAKQLTDADTKGSKSLSDLDSDCTLSILRLHILASGADVTSANAKRYOKRGGFDA 701

Db 956 TGYAKT---ALGEHLLNVGNRVNVEILQIF----- 985

Qy 702 DACMELRLSNPLVKKLSSTSVYDLTPGKRMKILHAL-----CGKLLTLVSTRDFIEDYVD 757

Db 986 ---MEAHGCGQELTESLTKFAQHTPAQKAVLAFLINELACSK--SVVSEIDKNIDYMS 1040

Qy 758 ILROAKQEFRELKAEQHKEREBAARIRKKEKLEKQEKQKKEKLEKEDFORNSTA 817

Db 1041 NLRDKNVV-----EGKLR-----KLRIHAKTKGKRTSG 1071

Qy 818 DISIGEEE-----REDFDTSIESKDEQKELDQDMFTEDDDPGSHKRRRG 864

Db 1072 GIDLGEEQHPLGTTPGKRRKRGSDYDDDDDDSDQDDEDEED-----KEDQKG 1127

Qy 865 KRQNGKFEFTROBQINCVTRELLTADEEE---ALKQEHQREKELLEKIQS----- 913

Db 1128 KKTD-----ICEDEDEGDAQAAVEELEKQTEKLSKQSQYRRKLFD 1168

Qy 914 AIACTNIFFPLGRDRMYRYWIFPSIPGLFIE--EDYSGLTEDMLLPSPSSQNNVQSODP 971

Db 1169 ASHSLRSVMFGPDYRRYRWILPCGGIFVEGMESGEGLEIAKEREKELKKAESVQIKEE 1228

Qy 972 QVSTKTGEPLMSESTSNIDQ-----PRDHSVOL 1000

Db 1229 MFET-SGDSLNCSTNDCEQKEDLKEKDNLTNLFLQKPGSFKLSKLEVAKMPPESEVMT 1287

Qy 1001 PKP-----VHK-----PWRWCFYSSCE 1017

Db 1288 PKPNAGANGCTLSYONGSKHSLGSVQSTATQSNVEKADNNLNTGSGGPK--FYSLPL 1345

Qy 1018 QLDOLIEALNSRGHRESAL-----KETLQEKSRICAOQLARFEEKHFHFDKQPD--- 1068

Db 1346 N-DOLLKTLTEKRWFSLLPRTPCDDTSLTHADMSTASLVTPOSQPPSKSPSTPAPLG 1404

Qy 1069 -----SKPTYSRG-----R 1077

Db 1405 SSAQNVPVGLNPFALSPLOVKGVSMMGLQFCGWPTGVVTSNIPTLSVPVSLGSLGLSEG 1464

Qy 1078 SSNAYDPSQWCAEQ-----LEL-RLRDF-----LLD 1103

Db 1465 NGNSFLTSNVASSKSESPVQNEKATSAQPAAVEAKPVPDFPSPKPIPEEMQFGWRIID 1524

Qy 1104 IEDRIYQGTLCAI-----KVTDRH---IMRSALSGRYELLS-BENKENG 1144

Db 1525 PED-----LKALLKVHLRGIREKALQKIQKHLDYITQACLKNKDVAILNENEHQ 1578

Qy 1145 IIKTV--NEDVEEMEIDEQTKVIVKRLLGIKTGTPTSTVSTNASTPOSVS----- 1192

Db 1579 VTRDIVENWSVEEQAMEMDLVLQOVDELRERRVASASLQVKGWMCPEPASEREDLVYFEH 1638

Qy 1193 -----SVVHYL-----AMALFOIEQOGERFLK--APLDA 1220

Db 1639 KSFTKLCKEHDGPTGEDESSAHLERKSDNPLDIAVTRLADLERNIERRIEADIAP--- 1695

Qy 1221 SDSGRSVKTVLDWRRESLSSASLSQVFLHSLTLDRSVWSKSTLNARCKICRKGGAEN 1280

Db 1596 -----GLRVWRALSEARNAQVLCIQLOKLSIAWEKSIWKYVCQICRKGDNEL 1746

Qy 1281 MVLCDGDRGHHTYCVRPKLTVPEDGWFCEPRKQ-----RCRRLSFRQPSLESDED 1335

Db 1747 LLLCDGCDKCHTYCHRPKITTPDGDWFCPACIAKASGQTLKIKLHVKGKTKNESKK- 1805

Qy 1336 VEDSMGDEDEVGDDEEGSEBEEYEVEQDEDDSDQEEVEVSLPKRGPQVRLPVKTRGK 1395

Db 1806 -----GKVTLTGDT-----DED----- 1819

Qy 1396 LSSFSRGGQOBPGRYPSRSQQSTPTKTVSSKTSRLRKINSAPPTETKSLRIASRSTR 1455

Db 1820 -----SASTSSSL----- 1827

Qy 1456 HSHGPLQADVVELLSPRKRRGRKSNANTPENSFPNFRVIAATKSESQSRSYNIASKL 1515

Db 1828 -----KRGKD-----LQKRKWEENTSIN---L 1847

Qy 1516 SLSESEKRRCKRQSPSPVTLGRSSRGQGVHLSAFQVLVLRHDDSWPFLK 1575

Db 1848 SKQESFTSVKPKRDDSK-----DLALCSMLTETMETHEDAWPFLP 1889

Qy 1576 VSKIQVPDYDIKKPIALNIIRKVNKCYKLASEFIDDIELMFSCFENPNTSEAK 1635

Db 1890 VNLKLVPGYKKVKKPMDFSTIREKLSGGQVPLETALDVRVLVFNCFETWEDDSIDGR 1949

Qy 1636 AGTRLQAFF 1644

Db 1950 AGHNMRYF 1958

RESULT 25

US-10-702-148-71

Sequence 71, Application US/10702148

Publication No. US20040063145A1


```

; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042002
; CURRENT APPLICATION NUMBER: US/10/702,148
; CURRENT FILING DATE: 2003-11-05
; PRIOR APPLICATION NUMBER: PRIOR APPLICATION NUMBER: US/09/839,479
; PRIOR FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 71
; LENGTH: 1969
; TYPE: PR1
; ORGANISM: Homo sapiens
; US-10-702-148-71

Query Match      8.2%; Score 704; DB 15; Length 1969;
Best Local Similarity 21.1%; Pred. No. 2.4e-30;
Matches 331; Conservative 196; Mismatches 426; Indels 616; Gaps 53;

QY 420 SPANRRGRPP-----KRIHSQEDNVA-----442
DB 662 SMRRRKGRPPNVGNABFDNADAKLLKQLQAEARQAQIKLRLKQKQEQARVAKAK 721
QY 443 -NKOTLASYSRKATKDKLQKE-----MKSLAFE-----KAKLREKADA--484
DB 722 KQATMAAEKREKQEQIKIHQKQEKIKRIQIRMEKELRAQOILEAKKKKEEANA 781
QY 485 LEAKKEKE-----DKEKREELUKKIVSEERLKKKEKELKVERKE---527
DB 782 LEAEKRIKEREMRROQAVLLKQERERRRRQHMLMKAMEARKKAEKELKQEKREKRL 841
QY 528 -REKLEBRKRYVEYLKQWSPREMECDLKLPE-PTPVKTRLPPIFGDALMVLFL 585
DB 842 NKERKLEQRLEWAKELKKNEDMLADQKPLPELPVGLVSGSTFSDCLMVQVFL 901
QY 586 NAFGEL--FDLQDEPFGVTLLEVEALVGN--DSEGPLCELLFFLTAIFAIAEIEEE 641
DB 902 RNFGKVLGFDVNIQVNP--LSVLQEGILLNIGDSMGVEQDILLVLLSA--AVCDPGLI 955
QY 642 VAKEQLTADTKGCSLKLSDLDSCITLSEILRHILASGADVTSANAKYRYKRGGFATD 701
DB 956 TGYKAKT---ALGEHLNVGNRDNVSEILQIF-----985
QY 702 DACMELRLSNPLSVKKLASTSVYDLTPGCKKILHAL-----CGKLLTLVSTDFIEDYVD 757
DB 986 ---MEAHCGQOTELTESLTKPAQANTPAQKAVLAFLINELACSK--SVSEIDKKNIDYMS 1040
QY 758 ILROAKQFRELKARHQRKEREAAARIRKKEELKQEQKQKQEKQKLEKDEORNGTA 817
DB 1041 NLRRDNVNV-----EGKLR-----KLRIHAKTKGRTSG 1071
QY 818 DISIGEE-----REDPDTIESKDTQEKELDDQDMFTFEDDDPGSHKGRRG 864
DB 1072 GIDLGEQHLPLGTPTPGRKRRKRGDSYDDDDDDSDQDQDEDEDEDE---KEDQKG 1127
QY 865 KRGQNGFKEFTQEQINCVTRRELLTADREE---ALKQEHQKKEKELLEKIQS-----913
DB 1128 KKTQD-----ICEDDEGQAAASVELEKQTEKLSKQSQYRRKLFD 1168
QY 914 AIACNIPPLGRDMRYRYPPSPGLFIE--EDYSGLTEDMLLPSPSSFNQNVQSDP 971
DB 1169 ASHSLRSVMFGPDRYRYRYWILPRCGGIFVEGMESGEGLIEIAKREKLKKAESVQIKEE 1228
QY 972 QVSTKGTGSLMSESTSNIDQG-----PRDHSVOL 1000

1229 MFET--SGDSLNCSTNDHCEQEDLKEKONTNLFLOKPGSFSKLSKLLVAKMPPSEVMT 1287
1001 PKP-----VHK-----PNCWCFYSSCE 1017
1288 KPNAGANGCTLSYNSGKHSGLSVQSTATQSNVNEKADSNNTLPTNGSSGFGK--FVSLP 1345
1018 QLDQLEALNSRGHRESAL-----KETLLOEKSRICAQARFSEBKPHFSKQPPD---1068
1346 N-DQLLKTLEKRWQFSLPRTPCDDTSLTHADNSTASLVTPQSQPPSKSPPTAPL 1404
1069 -----SKTYSRG-----R 1077
1405 SSAQNPVGLNPPALSPLOKVGVSMMGLQFCGWPTGVVTSNIPFTLSVPSLGLSG 1464
1078 SSNAYDPSCMAEQ-----LEL-RLRDF-----LLD 1103
1465 NGNSFLTNSVASSKSESVPQNEKATSAQPAAVEVAKPVPFPSPKPIPEMOMFGW 1524
1104 TEDRIYQGTGAI-----KVTDH-----IWSALESGRYBELL- EENKENG 1144
1525 PED-----LKALLKVLHLRGIREKALQKQIKHLDYITQACLKNKDVAIIELN 1578
1145 IIKTV--NEDVEEMEIDEQTKVIVKDRLLGKGTETPTSTVSTWASTPQSVS-----1192
1579 VTRDIVNSVEEQAMENDLSVLQQVEDLERRVASASLQVKGWMCPEPASEREDLY 1638
1193 -----SVVHYL-----AMALFOEQIEIRERFLK--APLDA 1220
1639 KSFTYKLCHEGDEFTGEDESSAHLERKSDNPLDJATVRLADLERNIERIEDIAP 1695
1221 SDSGRSYKTLDRWRRESLSSASQVFLHLSTLDRSVIWSKSIILNARCKICRKGDA 1280
1696 -----GLRVERRALSARSAAQVALCICQLOKSIAMEKSIMKVYQICRKGDN 1746
1281 MVLDCDGRGHHTYCVRPKLTVPBGDWFCECRPKQ-----RCRRLSFRQRPSP 1335
1747 LLLCDGCDKGCHTYCHRPKITTPDGDWFCPACIAKASGQTLKIKKLHVKGKKT 1805
1336 VEDSMGGDEDDVDGDEBEGQSEEBEVEQDEDDQSEEBEVSLEPKRGRPOVLPV 1395
1806 -----GKVTLTGTDE-----DED-----1819
1396 LSSFSRSGQOQEPGRYPSRQSQSTPTKTSVSSKTSRSLRKINSAPPTETKSLR 1455
1820 -----SASTSSSL-----1827
1456 HSHGPLQADVVELLSPRKRGRKRSKANSNTPNSPNFNRVIATKSEQSRSVNIAS 1515
1828 -----KRGND-----LQKRMKEENTSIN---L 1847
1516 SLOESKRCRCRKQSPSPVTLGRRSSGROGVHLSAFEQLVVELVRHDSWPLKL 1575
1848 SKQESFTSVKPKKDDSK-----DIALCSMILTEMETHEDAWPFLP 1889
1576 VSKIQVPDYDIIKKPIALNIIRKVNKCEYKLASEFIDDIELMFNSCFEYNPRTSE 1635
1890 VNLKLVPGYKVKIKPMDFTIREKLSGQYPNLETFALDVLVLDNCETFWEDSDIG 1949
1636 AGTRLOAFF 1644
1950 AGHNRKYF 1958

RESULT 26
US-10-087-192-666
; Sequence 666, Application US/10087192
; Publication No. US20020182586A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; APPLICANT: Engelhard, Eric K.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: CANCER
```



```

; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-839-479-70

Query Match
Best Local Similarity 7.9%; Score 682.5; DB 9; Length 1876;
Matches 311; Conservative 200; Mismatches 389; Indels 639; Gaps 50;

QY 424 RRRGPPRIHISQSDNVANKOTLASYSRKATKRDKLLKQEMSKLAFKAKLKR--EK 481
DB 644 RGRGRPPK-VKITELLNKTDRNPKLLEAQT-----LNEEDKAKIAKSKKQKQVOR 696

QY 482 ADAL-----BAKKEKEDAEK-----KREELK-KIVEERLUKKEE 516
DB 697 GECLTATIOQARNKPKQETSLKHKEAKKSKAEKERGKTQKELKEKVKREKKEVKKE 756

QY 517 KERL-----KVEEKEREKLEKRYKVEVLKQMSKPREDMCCDLKELPEPTVK-T 568
DB 757 KEEVTKAKPACAKADITLATORLEERQKQOMTLEEMKPTEDMCLTDHQPUPDFSRVPL 816

QY 569 RLPPFIQDALMVEFLNAFGBL--FDLQDEPPDGVTLEVEEALV-GNDSEGPLCELLF 625
DB 817 TLPSCAFSDCLTIVFLSPGVLFDPKQVP--SLGVLOEGLLCOGDSLGEVQDILV 873

QY 626 FFLTAI-----FOAIAEBEEVAKBQLTADTKGCSLKSOLDSDCTLSEILRLHLILASGA 680
DB 874 RLLKAALHDPGPPSYCQSKIL-----GEKVSEIPLTRDNVSEILRCFLMAYGV 922

QY 681 DVTSANAKRYQKRGFGFATDDACMELRLSNPLVKKLSSVSVDLTGCKMKIL----H 736
DB 923 E-----PALCDRLRTOPFOAQPFPQQAALVAFPVH 952

QY 737 ALCGKLLTVSTRDFIEDYVDILROAKOEFRELKAEQHKEREKAAAAIRKKEKLEQ 796
DB 953 ELNGSTL-----IINEIDKTESMSS--YRKKNWIVEGRRLRLKTVLAK-- 994

QY 797 EQMKKEKQKLEKEDQORNSTADISIGEBEREDFTSIESKOTEQKELQDQDFTEDDDPG 856
DB 995 -----RTGRSEVMGRPE-----ECLGRRRSRIMEETSGBEEBES 1034

QY 857 SHKRRGRKRG-----QNGKFEFTROEQINCVTRELLTADAEERALKQEHQKELKLEK 910
DB 1035 AAVPQRRGRDGEVDATASSIPELREQ-----IEKLSKRQUFFRKLKLLHS 1079

QY 911 IQSALACTNIFPLGRDMRYRWIIPPSIPGLFIEDYSGLTEDMLLP----- 957
DB 1080 SQMLRAVS-----LGQDRYRRYRWVLPYLAGIFVEG-----TEGNLVPEVVIKETTDSLKV 1130

QY 958 -----PQVS-----TKTGEF-----RPSFQNNVQSQD----- 970
DB 1131 AAHASLNPALFSMKMELAGSNTTASSPARARSRLTKPGFMQPREFKSPVRGQDSEQPQ 1190

QY 971 -----PQVS-----TKTGEF-----RPSFQNNVQSQD----- 980
DB 1191 AQLOPEAQLHVPAQPOPQLOLOLQSHKGFLQEGSPLSLGOSQHDLSQSAFLSMLSQTH 1250

QY 981 --LMSESTSNIDQGR-----DHSVQLP----- 1001
DB 1251 SLLSSSVLTPDSSPKGLDPAPSQPEPEPDEAESPDLOAFWNIQAQMPCNAAPTPP 1310

QY 1002 -----KPVHKPNR----- 1009
DB 1311 LAVSEDQTPPSQQLASSKPMNRPSAANPCSPVQFSSTPLAGLAPKRAGDPGEMPOSPT 1370

QY 1010 -----WCFYSSCEOLDQILIEALNSRG 1030
DB 1371 GLGQPKRRGRPPSKFFQMEQORVLTLTAQVPVPEMCSGWWIIPDPPEMLDAMLKALHPRG 1430

QY 1031 HRESALKETLLQEKSRICAQLARFEEKXFDKPDQPSKPTYSRGRSSNAVDPSCMAE 1090
DB 1431 IREKALHKLHNRDFLOEVCURPSADPI-----PEPQLOFAFOGQIMS--WSPKEKTYE 1483

QY 1091 KQLELRDLFDLIDIEDRIYQGTGLGAIKVTDRHWRSALESRYELLSEENKENGIIKTWN 1150

; Db 1484 --TDLAVLQWVEELEQRVIM-----SDLQIRGWTCPSPD-----STR 1518
QY 1151 EDVEEMEIDEOTKVIVKDRLLGIKTETPTSTVSTNASTPQSVSSVHYHVLAMALFOLEQSTE 1210
DB 1519 EDLAYCEHLSQSDIEDITWRGFGREGLAPQKTTNP-----LDLAVMLRLAALEQNVK 1569
QY 1211 RRFELKAPL-----DASDSGRSYKTV--LDRWRRESLSSASLSQV 1248
DB 1570 RRYLREPLWPTHMVELEKALLSTNGAPEGTTTTEISYEITPRIRIWRQTLQRCRCAAHVC 1629
QY 1249 LHLSTLDRSVIWSKILNARCKICRKGDAENWVLCDGCDRGHHTYCYRPKLKTYPEDGW 1308
DB 1630 LCLGHLERSIAWEKSVNVTCLVCRKGDNDFLLCDGCDRGCHTYCHRPKMEAVEPEGDW 1689
QY 1309 FCPECRPQRCRRLSFRQPSLESDVEDSMGGDDVDGDEEBEGSEEEVEVEQDED 1368
DB 1690 FCTVCLAQO-----VEGE----- 1702
QY 1369 DSQEBEEVSLPKRGRPQVRLPVKTRGKLSLSSFSRSGQOQEPYPSRQSQSTPKTTVASK 1428
DB 1703 -----FT-----QKPG-FPKRGQK-----RK 1717
QY 1429 TGRSLRKINSAPPTETKSLRTASRSTRHSHGFLQADVVELLSPRKR---RGRKSANNT 1485
DB 1718 SGYSL-----NFSEG-----DGRRRRVLLKGRES----- 1741
QY 1486 PENSPNPNFRVIATKSSSEQSRSVNIASKLSIQSESKRRCKRQSPSPVTLGRRSSG 1545
DB 1742 PAAGRYSEERL-----SPSKRR-----PUSMRNHS- 1768
QY 1546 RQGVHELSFAEQVLVLRHDDSWPFLKSVKIQVDPYDIKPKIALNIIREKVNKE 1605
DB 1769 -----DLTFCEIILMEMESHDAAPFLEPNPRLVSGYRRIIKNPMDFSTWRERLLRG 1822
QY 1606 YKLASEFIDDELMPSCNCFEYNPRNTSEAKAGTRLOAFF 1644
DB 1823 YTSSEFAADALLVDFNQCOTFNEDDSEVGKAGHIMRREF 1861

RESULT 28
US-10-376-537-71
; Sequence 71, Application US/10376537
; Publication No. US20030224405A1
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042001
; CURRENT APPLICATION NUMBER: US/10/376,537
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US/09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 71
; LENGTH: 1876
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-376-537-71

Query Match
Best Local Similarity 7.9%; Score 682.5; DB 15; Length 1876;
Matches 311; Conservative 200; Mismatches 389; Indels 639; Gaps 50;

QY 424 RRRGPPRIHISQSDNVANKOTLASYSRKATKRDKLLKQEMSKLAFKAKLKR--EK 481
DB 644 RGRGRPPK-VKITELLNKTDRNPKLLEAQT-----LNEEDKAKIAKSKKQKQVOR 696

QY 482 ADAL-----BAKKEKEDAEK-----KREELK-KIVEERLUKKEE 516
DB 697 GECLTATIOQARNKPKQETSLKHKEAKKSKAEKERGKTQKELKEKVKREKKEVKKE 756

QY 517 KERL-----KVEEKEREKLEKRYKVEVLKQMSKPREDMCCDLKELPEPTVK-T 568
DB 757 KEEVTKAKPACAKADITLATORLEERQKQOMTLEEMKPTEDMCLTDHQPUPDFSRVPL 816

QY 569 RLPPFIQDALMVEFLNAFGBL--FDLQDEPPDGVTLEVEEALV-GNDSEGPLCELLF 625
DB 817 TLPSCAFSDCLTIVFLSPGVLFDPKQVP--SLGVLOEGLLCOGDSLGEVQDILV 873

QY 626 FFLTAI-----FOAIAEBEEVAKBQLTADTKGCSLKSOLDSDCTLSEILRLHLILASGA 680
DB 874 RLLKAALHDPGPPSYCQSKIL-----GEKVSEIPLTRDNVSEILRCFLMAYGV 922

QY 681 DVTSANAKRYQKRGFGFATDDACMELRLSNPLVKKLSSVSVDLTGCKMKIL----H 736
DB 923 E-----PALCDRLRTOPFOAQPFPQQAALVAFPVH 952

QY 737 ALCGKLLTVSTRDFIEDYVDILROAKOEFRELKAEQHKEREKAAAAIRKKEKLEQ 796
DB 953 ELNGSTL-----IINEIDKTESMSS--YRKKNWIVEGRRLRLKTVLAK-- 994

QY 797 EQMKKEKQKLEKEDQORNSTADISIGEBEREDFTSIESKOTEQKELQDQDFTEDDDPG 856
DB 995 -----RTGRSEVMGRPE-----ECLGRRRSRIMEETSGBEEBES 1034

QY 857 SHKRRGRKRG-----QNGKFEFTROEQINCVTRELLTADAEERALKQEHQKELKLEK 910
DB 1035 AAVPQRRGRDGEVDATASSIPELREQ-----IEKLSKRQUFFRKLKLLHS 1079

QY 911 IQSALACTNIFPLGRDMRYRWIIPPSIPGLFIEDYSGLTEDMLLP----- 957
DB 1080 SQMLRAVS-----LGQDRYRRYRWVLPYLAGIFVEG-----TEGNLVPEVVIKETTDSLKV 1130

QY 958 -----PQVS-----TKTGEF-----RPSFQNNVQSQD----- 970
DB 1131 AAHASLNPALFSMKMELAGSNTTASSPARARSRLTKPGFMQPREFKSPVRGQDSEQPQ 1190

QY 971 -----PQVS-----TKTGEF-----RPSFQNNVQSQD----- 980
DB 1191 AQLOPEAQLHVPAQPOPQLOLOLQSHKGFLQEGSPLSLGOSQHDLSQSAFLSMLSQTH 1250

QY 981 --LMSESTSNIDQGR-----DHSVQLP----- 1001
DB 1251 SLLSSSVLTPDSSPKGLDPAPSQPEPEPDEAESPDLOAFWNIQAQMPCNAAPTPP 1310

QY 1002 -----KPVHKPNR----- 1009
DB 1311 LAVSEDQTPPSQQLASSKPMNRPSAANPCSPVQFSSTPLAGLAPKRAGDPGEMPOSPT 1370

QY 1010 -----WCFYSSCEOLDQILIEALNSRG 1030
DB 1371 GLGQPKRRGRPPSKFFQMEQORVLTLTAQVPVPEMCSGWWIIPDPPEMLDAMLKALHPRG 1430

QY 1031 HRESALKETLLQEKSRICAQLARFEEKXFDKPDQPSKPTYSRGRSSNAVDPSCMAE 1090
DB 1431 IREKALHKLHNRDFLOEVCURPSADPI-----PEPQLOFAFOGQIMS--WSPKEKTYE 1483

QY 1091 KQLELRDLFDLIDIEDRIYQGTGLGAIKVTDRHWRSALESRYELLSEENKENGIIKTWN 1150
```

Qy 482 ADAL-----EAKKKEKEDK-----KREELK-KIVEERLKKKEE 516
Db 697 GECLTTIOGQARNKFKQBTSLKKEAKKSKAKERKGTQKELKKEKREKKEKVKKE 756
Qy 517 KERL-----KVEREKEREKREKRYVYLKQWSPREDMECDLDELPEPTPVK-T 568
Db 757 KEEVTKAKPACKADKTLATQRLREERQKQMILEEMKKPTEDMCLTDHQPLDPSRVPGL 816
Qy 569 RLPEIFODALMWLEFLNAFEL---FDLQDFPQGVTLVELEBALV-GNDSEGLCELLF 625
Db 817 TLPAGFSDCLTIVEFLHSPGVLFDPKQVP---SLGVLOEGLLQCGDSLGEVQDILV 873
Qy 626 FFLTAI-----FOAIAEBEEVAXEQLTADTKGCSLKLSDLSCTLSEIILHLASGA 680
Db 874 RLLKAALHDPGPPSCQSKTL-----GEKVSEIPLTRDNVSEILURCLMAYGV 922
Qy 681 DVTSANAKRYQKRGFDATDACMELRLSNPSVLKKLSSTSVYDLTPGERMKIL----H 736
Db 923 E-----PALCDRLRTOPFOAQPPQQAALVAFPVH 952
Qy 737 ALCKGLTLVSTRFIEDYVDILROAKQEFRELKABQHRKEREBAARIRKREKKEKQB 796
Db 953 ELNGSTL-----IINEIDKTLSEMS--YRKNKWIVEGRLRLKTLAK-- 994
Qy 797 EOKMKQEKLEKEDQRNSTADISIGEEEREDFTSIESKDETEQKELDQDMFTEDDDPG 856
Db 995 -----RTGRSEVENGRPE-----ECLGRRSSRIMEETSMEEBEES 1034
Qy 857 SHKGRGRKRG-----ONGFKEFTREQINCVTRELLTADDEBALQEHQORKEKLEK 910
Db 1035 AAVGRRGRDGEVDATASSIPELREQ-----IEKLSKQLFFRKULLHS 1079
Qy 911 IQSAIACNIIPLGRDMRYRYWIPPSIPGLPIEDYSGLTEDMLLP----- 957
Db 1080 SQMLRAVS-----LQODRYRRRYWVLPYLAGFVFG-----TEGNLVPEEVIKKETDSLKV 1130
Qy 958 -----RPSSEQNVSOD----- 970
Db 1131 AAHASLNPALFSMKMELAGSNVTASSPARASRPLKTKPGMQPREFKSPVRGQDSQPQ 1190
Qy 971 -----POVS-----TKTGEP----- 980
Db 1191 AOLQPEALHVPAPQPOLQLOQSHKGLFQEGSPLSLGOSQHDLSQSAFLSWLSQTH 1250
Qy 981 --LMSSTNSIDQGR-----DHSVQLP----- 1001
Db 1251 SSLSSSVLTPDSSPGKLDPAPOPEPEPEDEAESPDLOAFWNISAQMFCAAPTTP 1310
Qy 1002 -----KPVHKNR----- 1009
Db 1311 LAVSEDQTPSPQOLASSKPMNRPSAANPCSVQFSPTPLAGLAPKRAGDPGEMPOSPT 1370
Qy 1010 -----WCFYSSCEOLDQLEBALNSRG 1030
Db 1371 GLGQKRGRRPPSKFFKQMEQORVLTOLTAQVPPEMCSGWWIIPDPEMLDAMLKALHPRG 1430
Qy 1031 HRESALKETLLOEKSRICAQIARLSEEFKHFSDKPDQPSKPTYSGRSSNAYDPSQMAE 1090
Db 1431 IREKALHKLHNRDFLQEVCLRPSADPI-----PEPRQLPAFOEGIMS--WSPKEKTYE 1483
Qy 1091 KQELRLRDLFLDIEDRIYQGTGNAIKYTDRIHRSALYESGRYELLSEENKENGIIKTVN 1150
Db 1484 --TDLAVLQWVEELEQVRIM-----SDLIQIGWTCPSPD-----STR 1518
Qy 1151 EDVEMEIDEQTKVIVKDRLLGKITETPSTVSTNASTPQSVSSVVHVLAMALFOEQIE 1210
Db 1519 EDLAYCEHLSQSDIITWRGPGREGLAQKRTNP-----LDLAVWELAALEQNVK 1569
Qy 1211 RRFUKAPI-----DASDSGRSYKTV--LDRWRRESLSSASLSQV 1248
Db 1570 RRYLREPLWPTHWVLEKALLSTNGAPEGTTTISYEITPRIIRWROTQRCRSHAVC 1629

Qy 1249 LHLSTLDRSVIWSKSIILNARCKICRKKGDAAENMVLCDGCDRGHHTYCVRPKLKTVPEGDW 1308
Db 1630 LCLGHLERSIANEWSVNVKVTCLVCRKGDNDFFLLCDGCDRGCHYCHRPKWEAVPEGDW 1689
Qy 1309 FCPCEPRQKRCRLRSFRQPSLESDEVEDSMGDEDEVDCGSEBEGOSEEYVEQDED 1368
Db 1690 FCTVCLAQ-----VEGE----- 1702
Qy 1369 DSQEBEVSUPLKRGQRPQVRLPVKTRGKLSSSFSSRGOQOEGRCYPSRSQQSTPKTTVSSK 1428
Db 1703 -----FT-----QKPG-FPKRGQK-----RK 1717
Qy 1429 TGRSLRKINSAPPTTKSLRIASRSTRHSHGLQADVVELLSPRRKE---RGRKSANNT 1485
Db 1718 SGYSL-----NFSEG-----DGRRRVLKGRES----- 1741
Qy 1486 PENSFPNPNFRVIATKSSEOSRSYNIASKLSQSESKRCKRQSPSPSVTLGRRSSG 1545
Db 1742 PAAGPRYSEERL-----SPSKRR-----PLSMRNHHS- 1768
Qy 1546 RQGVHSELSAFEQLVVELVRHDSWPFLKLVSKIQVDDYDIKKPIALNIIREKVNKE 1605
Db 1769 -----DLTFCEITLMEHSHDAAPFLEPVNPLVSGYRRIIKNPMDFSTMRELLRGG 1822
Qy 1606 YKLASEPIDDIELMFSNCFEYNPRNTSEAKAGTRLOAFF 1644
Db 1823 YTSSEFAADALLVFDNCQTFNEDDSEVGKAGHIMRRFF 1861

RESULT 29
US-10-702-148-70
; Sequence 70, Application US/10702148
; Publication No. US20040063145A1
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042002
; CURRENT APPLICATION NUMBER: US/10702,148
; CURRENT FILING DATE: 2003-11-05
; PRIOR APPLICATION NUMBER: PRIOR APPLICATION NUMBER: US/09/839,479
; PRIOR FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 70
; LENGTH: 1876
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-702-148-70

Query Match 7.9%; Score 682.5; DB 15; Length 1876;
Best Local Similarity 20.2%; Pred. No. 3.5e-29;
Matches 311; Conservative 200; Mismatches 389; Indels 639; Gaps 50;
Qy 424 RRRGRPPKRIHISOEDVANKQTLASYSRSKATKRDKLLKQEEEMKSLAFKAKLKR--EK 481
Db 644 RGRGRPPK-VKITELLNKTNRPLKLEAQET-----LNEEDKAKIAKSKKGRQKQV 696
Qy 482 ADAL-----EAKKKEKEDK-----KREELK-KIVEERLKKKEE 516
Db 697 GECLTTIOGQARNKFKQBTSLKKEAKKSKAKERKGTQKELKKEKREKKEKVKKE 756
Qy 517 KERL-----KVEREKEREKREKRYVYLKQWSPREDMECDLDELPEPTPVK-T 568
Db 757 KEEVTKAKPACKADKTLATQRLREERQKQMILEEMKKPTEDMCLTDHQPLDPSRVPGL 816

Qy 569 RLPPIFGDALMVLBFNAFGL---FDLQDFPDGVTLEVLBEALV-GNDSEGLCELLF 625
Db 817 TLPAGAFSDCLTIVFELSPGKVLGDFPAKQVP---SLGVLEGLCOGDSLGEVQDILLV 873
Qy 626 PFLTAI-----FOATAEEVEAKELTDADTKGCSLAKSLDLSCTLSEILRLHLTASGA 680
Db 874 RLLKAALHDPFPSPVCSQSKIL-----GEKVSEIPLTRDNVSEILRCFLMAYGV 922
Qy 681 DVTSANAKRYQKRGDFDADTACMELRLSNPLVKLSSTSVVDLTGPERMKIL-----H 736
Db 923 E-----PALCDRLRTQFPQAQPPQQAALVAPVH 952
Qy 737 ALCGKLLTLVSTRDFIEDYVILROAKQFRELKABQHRKEBEAAAIRKREKLEKEQ 796
Db 953 ELNGSTL-----IINEIDKTESMSS--YRKKNWIVGRLRLKTVLAK-- 994
Qy 797 EOKMKEKQELKEDEQRNSTADISIGEREDEDFTSIESKDEQELQDDMFTDEDDPG 856
Db 995 -----RGRSEVMGRPE-----ECLGRRRSRIMEETSGMEEEEEESI 1034
Qy 857 SHKRRGRKRG-----QNGKFETFRQPCINCVTRELLTADDEEALKQEHQKELILEK 910
Db 1035 AAVPGRRRGDEVDATASSIPELERQ-----IEKLSKQLFFRKULHS 1079
Qy 911 IQSAIACNIPPLGRDRMYRWIIPSPGLPIEBEDYSGLTEDMLLP----- 957
Db 1080 SQMLRAVS-----LGQDRYRRYVWLPYLAGIFVEG-----TEGNLVPEVVIKKTDSLKV 1130
Qy 958 -----RPSFQNNVQSOD----- 970
Db 1131 AAHASLNPALFSMKMELAGSNWTASSPARASRLTKPGMOPREFKSPVRGQDSEQP 1190
Qy 971 -----POVS-----TKTGEP----- 980
Db 1191 AOLQPEAQHVPAQPOFQLOLQSHKGFLEQSGSPLSLGOSQHDLSQSAFLSWLSQTH 1250
Qy 981 ---LMSSTSNIDQGR-----DHSVQLP----- 1001
Db 1251 SLLSSSVLTPDSSPGKLDPAPOPEPEPEAESSPDLOAFWNISAQMPCAAPTPP 1310
Qy 1002 -----KPVHKPNR----- 1009
Db 1311 LAVSEDQTPSPQOLASSKPMNRPSAANPCSPVQFSTPLAGLAPKRAGDPGEMPOSPT 1370
Qy 1010 -----WCFYSSCEOLDOLIEALNSRG 1030
Db 1371 GLGQPKRGRPPSKFFKQMEQORVLTLTAQVPVPEMCSGWIIIDPEMLDAMLKALHPRG 1430
Qy 1031 HRESALKETLLOEKSRI CAQLARFSEEEKFHPFSDKPPQDSKPTYGRGRSSNAYDPSQCAE 1090
Db 1431 IREKALHKLHNRDFLQEVCLRPSADPI-----PEPQLPAFQEGIMS--WSPKEKTYE 1483
Qy 1091 KQELRLRDLFDLIEDRIYQGTGLNAIKVTDRIHRSALSGRYELLSEENKENGIIKTVN 1150
Db 1484 ---TDLAVLQWVELEQRVIM-----SDLOIRGWTCPSPD-----STR 1518
Qy 1151 EDVEMEIDEQTKVIVKDRLLGIKTETPSTVSTNASTPQSVSSVVHYLAMALFOIEOGIE 1210
Db 1519 EDLAYCHELSDQEDITWRGFGREGLAPQRTKTHP-----LDLAVMLAALQNVK 1569
Qy 1211 RRFILKAPL-----DASDSGRSYKTV---LDRWRRESLSSASLSQVF 1248
Db 1570 RRYLREPLWPTHWELEKALLSTNGAPEGTTTISYEITPRIRWQTLQRCSAAHVC 1629
Qy 1249 LHLTLDRSVTWSKSIILNARKICRKKGDAENMVLCDGCDRGHTYCVRPKLKTVPEGDW 1308
Db 1630 LCLGHLESI AWKSVNKTCLVCRKGDNDLFLLLCDGCDRGCHYCHRPKWEAVPEGDW 1689
Qy 1309 FCPECRPQRCRLSFRQPSLESDDEDVDSMGGEDDDEVEGQSEEEYEVEQDED 1368
Db 1690 FCTVCLAQQ-----VEGE----- 1702
Qy 1369 DSQBEESVSLPKRGRPVRLPVKTRGKLLSSFSRGGQQQEPGRYPSRQQSTPKTVSSK 1428

Db 1703 -----FT-----OKFG-FPKRGOK-----RK 1717
Qy 1429 TGRSLRKINSAPPTETKSLRIASRSTRSHGFLQADVVELLSPRRKR---RGRKSANNT 1485
Db 1718 SGYSL-----NPSG-----DGRRRVLKGRS----- 1741
Qy 1486 PENSFPNFPNFRVIATKSQESRSVNIASKLSQSESKRRCKRQSPSPVTLGRRSSG 1545
Db 1742 PAAGFRYSEERL-----SPSKRR-----PLSMRNHHS- 1768
Qy 1546 RQGVHLSAPEQLVVELVRHDDSWPFLKLVSKIQVDPDYDIKKPIALNIIREKVNKE 1605
Db 1769 -----DLTFCETILMEMESHDAWPFLPVNPLVSGYRRIIKNPMDFSTMRERLLRG 1822
Qy 1606 YKLASEFTDIDELMFSCFENPNTSEAKAGTRLQAFF 1644
Db 1823 YTSSEFAADALLVFDNCQTFNEDDSEVGKAGHIMRRFF 1861

RESULT 30

US-10-087-192-663
; Sequence 663, Application US/10087192
; Publication No. US20020182586A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; APPLICANT: Engelhard, Eric K.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
; FILE OF INVENTION: CANCER
; FILE REFERENCE: 529452000122
; CURRENT APPLICATION NUMBER: US/10/087,192
; PRIOR FILING DATE: 2002-03-01
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 2059
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 663
; LENGTH: 1586
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-087-192-663

Query Match 7.9%; Score 680; DB 13; Length 1586;

Best Local Similarity 21.8%; Pred. No. 3.9e-29;
Matches 295; Conservative 178; Mismatches 387; Indels 492; Gaps 47;

Qy 424 RRRGRPPKRIHISQEDNVANKQTLASYSRKATKRDKLLKOEEMKSLAFKAKLKREKAD 483
Db 581 RGRGRPFK-----IKMPELL-----NKTQNLPPKLETOELLEDKAKMTK-- 622
Qy 484 ALEAKKEKEDKREKREKREKREKREKREKREKREKREKREKREKREKREKREKREKREK 542
Db 623 ---NKKMRQKVGQSGSPVQVQARNRKQDTKSLKQKQDKKLLKRLLEEQQRQAIL 678
Qy 543 KOWSKPREDECDKDLKLPPTPVK--TRLPEIFGDALMVLFLNAFGL--FDLQDFP 599
Db 679 EEMKPTEDMCLSDHQPLPDPFTRIPGLTSSRAFSDCLTIVEFLHSGKVGFLDTKQVP 738
Qy 600 DGVTVLEBALV-GNDSEGLCELLFFLTAIFAIAEABEEVEAKELQTDADTKGCSLK 658
Db 739 ---SLGVLEGLLCOGDSLQVQDLVRLKAL-----HDPGLPPYCSQLKILGEKWS 789
Qy 659 SLDDSDCTSLRILHLILHLAGADVTANAKRYQKRGDFDADTACMELRLSNPLSVKLL 718
Db 790 EIPLTRDNVSEILRCFLMA-----YRVE-----PSFCDL 819
Qy 719 SSTSVYDLTPEKMKILHALGCLLTLVSTRDFTDYVILROAKQEFRELK--AEQHRK 776
Db 820 RTQBFQAPPOQKAAIAFLVHEL-----NSTIILNIDKTLSESVSSCRKNKWIVEGLR 875
Qy 777 EREBAARIRKREKLEKQEQKM-KEQKQELKEDEQRNSTADISIGEREDEDFTSIES 835

Db 876 RLKTAALAKRTGRPEVMGAEADGLRRSSRIMEETS-----GIBEEBENTTAV-- 926
Qy 836 KDTQKELDDQMFTEDEDDPSSHKGRGRKGQNGFKFTQEQINCVTRELLTADSE-E 894
Db 927 -----HRRGR-----KEGIDVAASIPELERHIE 952
Qy 895 ALKQEHORKEKELLEKIQSAIACNIPPLGRDRMYRWIIPSPGLFIE--EDVSG-- 949
Db 953 KLSKRLQFFRKLHLSSQMLRAVS-----LGQDRYRWIYVLPYLAGIFVGESEGSTGWP 1008
Qy 950 -----LTEDMLLP-----RPSSFQNNVQSDQPVSTKT-----GEP 980
Db 1009 NFSAIQPCDAAPTTPPAVSEDOPTSLQLLASSKPMNTPGAANPCSPVQLSSHTLPGT 1068
Qy 981 L-----MSESTNIDQGR-----DHSVQL-----PKPVHKNRWCYSS 1015
Db 1069 KRLSGDSEEMSQPTGLQPKRRGRPPSKFKQVEQHLYLTQTAQPIPPMCMGMMWIRD 1128
Qy 1016 CEQLDOLIEALNSGRHESALKETLLQEK-----SRICAQLARFSEKHFSDKPODPSKPT 1072
Db 1129 PETUDVLKALHPRIKREKAUHKHLSKHQFLQEVCL-----QPLTDP- 1171
Qy 1073 YSRGRSSNAYDPSQWCAEKQLELRDLFLLDIEDRIYQGTILGAIKVTDRIHWSALESGR 1132
Db 1172 -----IPEPNEL-----PALBEG- 1184
Qy 1133 YELLSBENKENGIIKTWNED-----VEEMIDQTKVIVKD--RLLGKTKETPTVSTNA 1185
Db 1185 --VMSWSPKE---KTYETDLAVLQWVEELF---QRVLSLDLQIRGWTCPDPTDSTREDL 1234
Qy 1186 S-----TPQSV-----SSVHYLAMALFQIEQGLERELFKAPL--- 1218
Db 1235 TYCEHLDPSPEDIPWRGGRREGTVPQONNPLDLAVMLAVLEQNVERRYLRPLWAH 1294
Qy 1219 -----DASDSGRSYKTV--LDRWRESILSSASISQVFLHLSTLDRSVI 1260
Db 1295 EVVVEKALLSTPGAPDGTSTEISYEITPRVVRVWRTILERCSSAAQVCLNGQLERSIAW 1354
Qy 1261 SKSILNARCKICRKGDAENNVLCGDGRGHHTYCVRPKLTVPPEGWFCPECPKORCR 1320
Db 1355 EKSYNKVTCLVCRKGDNDEFLLCDGCDRGCHIYCHRPKMEAVPEGWFCVAVCL----- 1408
Qy 1321 RLSFRQPSLESDEVEDSGMGDEVDGDEEEQSEEEVEVEQDEDDQOEEREEVSLPK 1380
Db 1409 -----SQQVEEYR--- 1416
Qy 1381 RGRPVRLPVKTRGLKSSFSRQOQEPGRYPSPRSQQSTPKTVSSKTRSLRKINSAP 1440
Db 1417 -----QRPQ-FPKRGQK----- 1427
Qy 1441 PTETKSLRIASRSTRSHGHPLOADVVELLSPRRRRGRKSNANTPENSFPNFRVIAT 1500
Db 1428 -----RKS-----SFP----- 1433
Qy 1501 KSSEQRSVNIASKLSIQESESKEB--CRKQS-----PEPSPVTIGRSSRGQGVHE 1552
Db 1434 -----LTTFEGDSRRLMSRGRSDSPAVPRYPEDGSLSPKRRRHSHSRSHSD 1479
Qy 1553 LSAPEQLVVLVRHDSWPFLKYSKIQVPDYDIKKPIALNIREKNVKNCEYKLASEF 1612
Db 1480 LTFCEIILMEMESHDAAWPFLEPNRPLVGYRVRVKNPMDFFSTWRERLLGGTYSSEF 1539
Qy 1613 IDDIELMFSNCFEYNPRINTSEAKAGTRLQAPF 1644
Db 1540 AADALLVDNQCOTENEDDSEVGKAGHVNRFF 1571

RESULT 31
US-09-839-479-13
; Sequence 13, Application US/09839479
; Publication No. US2002003979A1
; GENERAL INFORMATION:

; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042002
; CURRENT APPLICATION NUMBER: US/09/839,479
; CURRENT FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 1878
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)..(1878)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-09-839-479-13

Query Match 7.8%; Score 671.5; DB 9; Length 1878;
Best Local Similarity 19.8%; Pred. No. 1.4e-28;
Matches 305; Conservative 197; Mismatches 399; Indels 639; Gaps 44;

Qy 424 RRRGRPPK-RIH--ISOEDN-----VANKQTLASYSRSKATKRDKLLKQEMKSL 470
Db 644 RGRGRPPKVIITELLNKNTPRLKCLAEQETLNEEDKAKIAKSKKKVQKQVQCECLTTI 703
Qy 471 AFEKAKLRKADAL---EAKKKEKEDKEK---KREELKKIVBER---RLKKEE--- 516
Db 704 QGQARNRKQETKSLHKEAKKLSXAEKGTQKQKLEKVKREKKEKVKMKEKEEVTK 763
Qy 517 -KEELKVEREKEREKLEERKRYVEYLKQNSKPRDEMCDDKELPEPTPK-TRLPPEI 574
Db 764 AKPACKADKTLATQRRLEERQKQOMILEMKKPTEDMCLTDHQPDPFSRVPGLTLPSCGA 823
Qy 575 FGDAVMVLEFLNAGFEL--FDLQDEFDPDGTVELEBALV-GNDSEGPLCELLPFFLTAI 631
Db 824 FSDCLTIVELHSGFQVGLGDPKADVP---SLGVLOEGLICQGDLSLGEVQDILLVLLKAA 880
Qy 632 F-----QATAEEEEEVAKQOLTADTKGCSLSKSLDSDCTLSILRLHILIASADV 682
Db 881 LHDPPSPYQSLKILGEKVSEIPLTRDN-----VSEILRCFLMAYGV-- 923
Qy 683 TSANAKYRYQKRGGFDDATDDACMELRLSNPSLVKCLKSSTSVYDLTPGKMKIL---HAL 738
Db 924 -----XPALCDRLRTQFPQAPQPOOKAAVLAPFVHEL 955
Qy 739 CGKLLTLVSTRDFIEDYVDILROAKQEFRLKAEQHRKEREAAARIRKKEEKLKEQE 798
Db 956 NGSTL-----IINEIDKLTESKSS--YRKNWIVEGRRLRLKTVLAK----- 995
Qy 799 KMKEQKLEKEDBQNSTADISIGEEREDPDTIESKOTEQKELDDQMFTEDEDDPGSH 858
Db 996 -----RTGRSEVEMGRPE---ECLGRRRSRRIIMEETSMEBEEESIAA 1037
Qy 859 KRGRGRKRG-----QNGFKFTQEQINCVTRELLTADSEALKQEHORKEKELLEKIQ 912
Db 1038 VPGRRGRRDGEVDATASSIPELERQ-----TEKLSKROLFRKKLLHSSQ 1082
Qy 913 SAIACTNIFPLGRDRMYRWIIPSPGLFIEDYSGLTEDMLLP----- 957
Db 1083 MLRAVS-----LGQDRYRWIYVLPYLAGIFVEG-----TEGNLVPVEVIKKETDSLKVA 1133
Qy 958 -----RPSSFQNNVQSD----- 970
Db 1134 HASINPALFSKMKMELAGSNNTTASSPARARSLPKTKPGFMQPRHFKSPVRGQDSEQPAQ 1193


```
QY 971 -----PQV-----STKTGE 979
DQ 1194 LQPEAQLHVPAPQPOLQLOLQSHKGFLQEQSGPLSLQSOHDLQSQAFSLWSLQTSQS 1253
QY 980 PLMSSESTNIDQGR-----DHSVQLP-----1001
DQ 1254 SLLSSSVLTDPSSFGKLPAPSPQPEEPDEAESPDLOAFWFNISQAFMPCNAAPTPL 1313
QY 1002 -----KPVHKPNR-----1009
DQ 1314 AVSEDOPTSPQOLASSKPMNRPSAANPCSPVQFSSTPLAGLAPKRRAGDPGEMPQSPG 1373
QY 1010 -----WCFYSSCEQLDOLIEALNSRGH 1031
DQ 1374 LGQPKRRGRPPSKFFQMEQRYLTQLTQAQVPPPCSCGMWIPDPPEMLDMLKALHPRGI 1433
QY 1032 RESALKETLLQEKSRICQAARFSEKPHFSDKPODKPTVSRGSSNAYDPSQMCABK 1091
DQ 1434 REKALHKLHNRDPLQEVCLPSADPIF-----EPRQLPAFQEGIMS--WSPKEXTYE- 1485
QY 1092 QLELRDLPLDIEDRIYQGTGALIKVTRHWRSALESGRYELLSEENKENGIIKTVNE 1151
DQ 1486 -TDLAVLQWVELEQRVIM-----SDLQIRGWTCPSPD-----STRE 1521
QY 1152 DVEEMEIDQTKVIVKDRLLGLTKTPTSTVSTNASTPOSVSVVHYLAMALFOIBOGIER 1211
DQ 1522 DLAYCEHLSDSQEDITWRGPGREGIAPQRTTNP-----LDLAVMLAALFQNVKR 1572
QY 1212 RFLKAPLSDSGRYSKYTVLDR-----WRESLSSASL 1244
DQ 1573 RYLRPL-----WPTHEVLEKALLSTNGAPEGTTTETISVEITPRIIRWQTLOQRCSA 1627
QY 1245 SOVFLHLSTLDRSVTWSKILNARCKIRKKGDAENMVLCDGCDRGHTYCVRPKLKTVP 1304
DQ 1628 AHVCLGLHLERSIAWEKSVNKVCLVCRKGNDFLLLLCDGCDRGCHYCHRPKEAVP 1687
QY 1305 EGDWFCPCRPKQRCRRLSFRQRPSSLEDEVEDSMGGEDEVDGDEEGQSEEEVEYE 1364
DQ 1688 EGDWFCVCLAQ-----VEGEFTQKPGFKRGQKRKSGYSLN 1725
QY 1365 QEDDSQEEVEVSLPKRGPQVRLPVKTRGKLSFSSFGQOQEGRPVPSRQOSTPKTT 1424
DQ 1726 FSEG-----GR-----RRVLLKGR-----ESPAAGPRYSE-----1752
QY 1425 VSSKTRSLRKINSAPPTETKSLRIASRTHSHGLQADVVELLSPRKRRRGRKSANN 1484
DQ 1753 -----ERLSKRR-----1762
QY 1485 TPENSPNPNFRVIATKSSQSRSVNIASKLSQESKRRKRKQSPSPVTLGRSS 1544
DQ 1763 -----LSMRNHS-----1770
QY 1545 GRQGVHLSAPEQLVVLVRHDDSWPFLKLSKIQVPDYDIIKKPIALMIIRKVNKC 1604
DQ 1771 -----DLTFCETILMEMESHDAWPFEPYVNLVSGYRIIKNPNDFSTMRLLRG 1823
QY 1605 EYKLASEFIDDLIELMFSNCFEYNPNTSEAKAGTRLOAFF 1644
DQ 1824 GYTSSEFAADALLVFDNCQTFNEDDSEVGKAGHIMRRFF 1863
```

RESULT 32

```
US-10-376-537-13
; Sequence13, Application US/10376537
; Publication No. US20030224405A1
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042001
; CURRENT APPLICATION NUMBER: US/10/376,537
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US/09/418,710
; PRIOR FILING DATE: 1999-10-15
```

```
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 1878
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(1878)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-10-376-537-13

Query Match 7.8%; Score 671.5; DB 15; Length 1878;
Best Local Similarity 19.8%; Pred. No. 1.4e-28;
Matches 305; Conservative 197; Mismatches 399; Indels 639; Gaps 44;

QY 424 RRRGRPPK-RTH--ISOEDN-----VANKQTLASYSRKATKRDKLLKQEMKSL 470
DQ 644 RGRGRPPKVKITELINKTDNRPLKLEAQETLNEEDKAKIAKSKKMRQKVQRGECLTTI 703
QY 471 AFEKAKLKREKADAL---EAKKKEKDEK---KREELKKIVEEE-----RLKKEE--- 516
DQ 704 QGQARNKXQETKSLKHKEAKKSAEKGKTKQEKLEKVKREKKEKVKMKEEVTK 763
QY 517 -KELKVEREKERKLEKRYVEYLKQWSKPREDECDLDELPEPTPVK-TRLPEI 574
DQ 764 AKPACAKDKTLATQRLLEERQKQMILEEMKKPTDMCLTDHQLPDRFVRPGLTSPGA 823
QY 575 FGDALMVLFLNAGEL--FDLQDEFDGVTVLELEALV-GNDSEGPLCELLPFFLTAI 631
DQ 824 FSDCLTIVFLLHSGKVLGFPDPAKVP---SLGVLEGLLCQGSOLGVEQDLVLLKAA 880
QY 632 F-----QAIAEEEEVAKEQLTDADTKGCSLSLDLDSCTLSEILRLHILASGADV 682
DQ 881 LHDPGFPYSQSLKILGEKVSEIPLTRDN-----VSEILRCFLMAYGV-- 923
QY 683 TSANAKYRYQKRGDFDADTACMBELRLSNPSLVKLSSTSVYDLTPGEMKIL-----HAL 738
DQ 924 -----XPALCDRLRTQPPQAPQPPQAKAALVAFPPVHEL 955
QY 739 CGKLLTLVSTDFIEDYVILRQAKQFRELKQHKREKEREAAAARIRKKEKLEKEQEQ 798
DQ 956 NGSTL-----IINEIDKLTESMSS--YRKNKWIIVEGRRLRLKTVLAK----- 995
QY 799 KMKEKQEKLEDEQRNSTADISIGEEREDFDTSIESKDTQEKELDQDMFTEDDDPGSH 858
DQ 996 -----RTGSEVEMGRPE-----ECIGRRSSRIMEETSQMEEEEEESIAA 1037
QY 859 KRGRGRKRG-----QNGFKFEFTQEQINCVTRELLTADDEEALKQEHQKELLEKIQ 912
DQ 1038 VPGRRGRREDGEVDATASSIPELERQ-----IEKLSKRLQFFRKLLHSSQ 1082
QY 913 SAIACTNIFPLGRDRMRYRYWFFSIPGLFTEEDYSGLTEDMLLP----- 957
DQ 1083 MLRAVS-----LGODRYRRRYWVLPYLAGIFVEG-----TEGNLVPEEVIKKTDSLKVA 1133
QY 958 -----RPSGFQNNVQSQD----- 970
DQ 1134 HASLNALFSMKMELAGSNNTTASSPARARSRLTKPGFMQPRHFKSPVRGQDSQEQPQAQ 1193
QY 971 -----PQV-----STKTGE 979
DQ 1194 LQPEAQLHVPAPQPOLQLOLQSHKGFLQEQSGPLSLQSOHDLQSQAFSLWSLQTSQS 1253
QY 980 PLMSSESTNIDQGR-----DHSVQLP-----1001
DQ 1254 SLLSSSVLTDPSSFGKLPAPSPQPEEPDEAESPDLOAFWFNISQAFMPCNAAPTPL 1313
```

```
Qy 1002 -----KPVHKPNR----- 1009
Db 1314 AVSEDOPTSPQQLASSKPMRPSAANPCSPVQFSSTPLAGLAKRRAGDPGEMPQSPGTG 1373
Qy 1010 -----WCFYSSCEQLDQLIEALNSRGH 1031
Db 1374 LGQPKRRGRPPSKFKQMEQRYLTQLTAQPVPPMPCSGMMWIPDPEMLDAMLKALHPRGI 1433
Qy 1032 RESALKETILQEKSRICAOARFSEKHFSDKQPDSPKTYSGRGSNAYDPSPQMAEK 1091
Db 1434 REKALHKLHNRHDFLOEVLCPRSADPFI-----EPRQLPAFOEGIMS--WSPKEKTYE- 1485
Qy 1092 QLEURLRDFLLDIDRIYQGLGAIKVTDRHIMSALESGRYELLSEENKENGIIKTVNE 1151
Db 1486 -TDLAVLQWVEELQVRIM-----SDIQIGWTCPSDP-----STRE 1521
Qy 1152 DVEEMEIDEQTKIVKDRLLGIKTTPTSTVNASTPQSVSVHYLAMALFOEQGIER 1211
Db 1522 DLAYCEHLSDSQEDITWRGPGREGLAQORKTNP-----LDLAVMLAALQNVKR 1572
Qy 1212 RFLKAPLDASGRSYKTVLDR-----WRESLLSASL 1244
Db 1573 RYLREPL-----WPTHEVLEKALLSTPNGAPEGTTTSEIYTPRIRIWRQTLQRCRSA 1627
Qy 1245 SOVFLHSLTDRSVIWSKSILNARCKICRKGDAENMVLCDGCDRGHTYCVRPKLATVP 1304
Db 1628 AHVCLCLGHLSRTAWKSVNKTCLVCRKGDNDDEFLLCCDGRGCHYCHRPMEAVP 1687
Qy 1305 EGDWFCPECPKQRCRLSFRQPSLESDEVDMSGDEDDVDGDEEGQSEBEEYEVE 1364
Db 1688 EGDWFTVCVLAQ-----VEGEFTQKPGFPKRGQKRSGYSLN 1725
Qy 1365 QDEDSQEBEEVLPKGRQVRLPVKTRGKLSSFSRGGQQPGPYPRSRQOSTPKTT 1424
Db 1726 FSEGD-----GR--RRVLLKGR-----ESPAAGPRYSE----- 1752
Qy 1425 VSSGTGRSLRKINSAPTETKSLIASRSTRHSHGPLQADVVELLSPRKCRGRKSANN 1484
Db 1753 -----BRLSPSKRRR----- 1762
Qy 1485 TPENSPNFPNFRVIATKSSQSRSVNIASKLSQESKRRCKRQSPSPVTLGRRSS 1544
Db 1763 -----LSMENHHS----- 1770
Qy 1545 GRQGVHELGAFOQLVVELVHRDSDWPLKLVSKIQVDPYDIYIKKPIALNIIREKNKC 1604
Db 1771 -----DLTFCEILMEMESHDAWPFXEPVNPRLVSGYRRRIIKNPMDFSTWRLLRG 1823
Qy 1605 EYKLASEFIDDIELMFNCFEYNPRNTSEAKAGTRLQAFF 1644
Db 1824 GYTSSEFAADALLVFNQCTFNEDDSEVGKAGHIMRRFF 1863
```

RESULT 33

```
US-10-702-148-13
; Sequence 13, Application US/10702148
; Publication No. US20040063145A1
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042002
; CURRENT APPLICATION NUMBER: US/10/702,148
; CURRENT FILING DATE: 2003-11-05
; PRIOR APPLICATION NUMBER: PRIOR APPLICATION NUMBER: US/09/839,479
; PRIOR FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
```

```
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 1878
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(1878)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-10-702-148-13
```

```
Query Match 7.88; Score 671.5; DB 15; Length 1878;
Best Local Similarity 19.88; Pred. No. 1.4e-28;
Matches 308; Conservative 197; Mismatches 399; Indels 639; Gaps 44;

Qy 424 RRRGRPPK-RIH--ISOEDN-----VANKQTTLASYSRKATKRDKLLKQEEKMSL 470
Db 644 RGRGRPPKVIITELLNKTDNRPLKLEAQETLNEEDKAKIAKSKKKMRQKVQVQGECLTTI 703
Qy 471 AFKAKLKRKADAL---EAKKKEKEDKEK---KREELKIVVEE-----RLKKEE--- 516
Db 704 QGOARNKRKQETKSLKHKEAKKSKXAEKRGKTQKQELKEKVKREKKEKVKMKKESEBVTK 763
Qy 517 -KERLKVREKEREKREKRYVEYLKOWSKPREDMECDDLKELPEPTPVK-TRLPEEI 574
Db 764 AKPACKADKTLATQRLLEERQKQOMILEEMKKTEDMCLTDHQLPDPFSRVPGLTLPSCGA 823
Qy 575 FGDALMVLFLNAGEL--FDLQDEFPDGVTVLEVEALV-GNDSEGPLCELLFFFTLTAI 631
Db 824 FSDCLTIVFLHSGFVKVLGDFDPAKDPV---SLGVLQEGLLCQGDSLGEVQDILLVRLKAA 880
Qy 632 F-----QAIABEEVEVAKQELTDADTKGSGLSKSLDSDCTLSILRLHILASADV 682
Db 881 LHDPGFSPYCSQSLKILGEKVSEIPLTRDN-----VSBILRCFLMAYGV-- 923
Qy 683 TSANAKRYQKRGFDATDDACMELRLSNPLVKLSSTSVYDLTQCEKMKIL----HAL 738
Db 924 -----XPALCDRLTQFPQAPPOQKAALVAFVHEL 955
Qy 739 CGKLLTLVTRDFIEDYDILROAKQFRELKAEQHRKEREAEAAARIKRKEEKLKEQEQ 798
Db 956 NGSTL-----IINEIDKLTLESMS--YRKNKWIIVEGRLRLKTLAK--- 995
Qy 799 KMKQEKLEKEDQORNSTADISGEHEREDFDTSIESKQTEQKELDQDMTEDEDDPGSH 858
Db 996 -----RTGRSEVENGRPE-----ECLGRRRRSSRIIMEETSGMEEEEEESIAA 1037
Qy 859 KRGRGRKRG-----QNGFKFTQEQINCVTRELLTADDEEALKQSHQKKEKELLEKIQ 912
Db 1038 VPGRRGRDGEVDATASSIPELERQ-----IEKLSKQLFRKKLLHSSQ 1082
Qy 913 SAIACTNIPFLGRDRMYRYWFIPIFGLFIEDYSGLTEDMLLP----- 957
Db 1083 MLRAVS-----LGDRYERYRWVLPYLAGIIFVEG-----TEGNLVPEEVIKKTDSLKAA 1133
Qy 958 -----RPSSFQNNVQSOD----- 970
Db 1134 HASINPALFSKMKMELAGSNTTASSPARARSRLKTPGMQPRHFKSPVRGQDSEQPOAQ 1193
Qy 971 -----POV-----STKTGE 979
Db 1194 LQPEAQLHVPAQPOPQLQLOLQSHKGLFEOGSGPLSLGQSQHDLSSQAFSLWSLQTSUHS 1253
Qy 980 PLMSESTSNIDQGR-----DHSVQLP----- 1001
Db 1254 SLLSSSVLTPDSSPGKLDPAPOPEPEPEDEAESSPDLQAFWNISAQMPCAAPTPL 1313
Qy 1002 -----KPVHKPNR----- 1009
Db 1314 AVSEDOPTSPQQLASSKPMRPSAANPCSPVQFSSTPLAGLAKRRAGDPGEMPQSPGTG 1373
```

QY 1010 -----WCFYSSCEQLDQLIALNSRGH 1031
Db 1374 LGQPKRGRPPSKFKQMEQRYLTQTAQVPPPEMCSGMMWIPDPEDMDLAKLHPRG1 1433
QY 1032 RESALKETLLQEKSRICAQLARFSEKHFSDKPODPTVSRGRSSNAYDPDSQMAEK 1091
Db 1434 REKALHKLHNRDLQEVCLRPSADPIF-----EPQIPAFQEGIMS--WSPREKTYE- 1485
QY 1092 QLELRDLFLDIEDRIYOGTLGAIKVTDRHIWRSALSGRYELLSEENKENGIIKTVNE 1151
Db 1486 -TDLAVLQWVEEQRVIM-----SDLQIRGWTCPSPD-----STRE 1521
QY 1152 DVEEMEIDEQTKVIKVDLLGTLGKTETPTSTVNASTPQSVSVHYLAMALFOIEQIER 1211
Db 1522 DLAYCEHLSQOEDITWGPREGIAPQRTNP-----LDLAVMLAALQVQYKR 1572
QY 1212 RFLKAPLSDSGRSYKTVLDR-----WRESILLSSASL 1244
Db 1573 RYLREPL-----WPHVEVLEKALLSTNGAPEGTTTISVEITPRIRWQTLQCRSA 1627
QY 1245 SOVFLHLSTLDRSVIWSKSIILNARCKICRKGDAENNVLCDCGRGHHTYCVRPKLKTVP 1304
Db 1628 AHVCLCLGHLERSIAWKSIVNKVCLVCKRGDNDEFLLLDCDCRGCHYCHRPKMEAVP 1687
QY 1305 EGDWFCPCPKQRCRRLSFRQRPSEDEDEVDGDEEGQSEEEVEVE 1364
Db 1688 EGDWFCVCLAQ-----VEGEFTQKPGFRGQKRGSGYSLN 1725
QY 1365 QDEDSQSEEEVSLPKRGPQVRLPVKTRGKLSSFSRGGQOQEPGRYPSRQOSTPKTT 1424
Db 1726 FSEGD-----GR-----RRVLKGR-----ESPAAGRYSE----- 1752
QY 1425 VSSGTRSLRKINSAPPTETKSLRIASRSTRHSHGLQADVVELLSPRKRRGRKKSANN 1484
Db 1753 -----EELSPSKRR----- 1762
QY 1485 TPENSPNPNFRVIATKSSQSRSVNIASKLSQSESKRRCKRQSPSPVTLGRSS 1544
Db 1763 -----LSMRNHS----- 1770
QY 1545 GQGGVHELSAFEQVLVLRHDSWPFLKLVSKIQVDPYDIKKPIALNIREKVNKC 1604
Db 1771 -----DLTCEILMESHDAWPPXEPVNPRLVSGYRIIKNPMDFSTMRELRUG 1823
QY 1605 EYKLASEFIDIELMFSCFYNPNTSEAKAGTRLOAFF 1644
Db 1824 GYTSSEEFADALLVFDNCQTFNEDDSEVGKAGHIMREFF 1863

RESULT 34
US-10-195-730-317
; Sequence 317, Application US/10195730
; Publication No. US2003014492A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et. al
; TITLE OF INVENTION: 101 Human Secreted Proteins
; FILE REFERENCE: P2017P1
; CURRENT APPLICATION NUMBER: US/10/195,730
; CURRENT FILING DATE: 2002-07-16
; PRIOR APPLICATION NUMBER: US/09/281,976
; PRIOR FILING DATE: 1999-03-31
; PRIOR APPLICATION NUMBER: 60/060,837
; PRIOR FILING DATE: 1997-10-02
; PRIOR APPLICATION NUMBER: 60/060,862
; PRIOR FILING DATE: 1997-10-02
; NUMBER OF SEQ ID NOS: 390
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 317
; LENGTH: 149
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (128)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (133)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (136)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

; LOCATION: (128)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (133)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (136)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (140)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (143)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-195-730-317
Query Match 7.7%; Score 662; DB 14; Length 149;
Best Local Similarity 96.1%; Pred. No. 2e-29;
Matches 122; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 1250 HLSTLDRSVIWSKSIILNARCKICRKGDAENNVLCDCGRGHHTYCVRPKLKTVPEGDW 1309
Db 12 HLSTLDRSVIWSKSIILNARCKICRKGDAENNVLCDCGRGHHTYCVRPKLKTVPEGDW 71
QY 1310 CPECPRKORCLFRQRPSEDEDEVDGDEEGQSEEEVEVEQDEDD 1369
Db 72 CPECPRKORCLFRQRPSEDEDEVDGDEEGQSEEEVEVEQDEDD 131
QY 1370 SQEEVEV 1376
Db 132 SXEEVEV 138
RESULT 35
US-10-799-747-317
; Sequence 317, Application US/10799747
; Publication No. US20040157258A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et. al
; TITLE OF INVENTION: 101 Human Secreted Proteins
; FILE REFERENCE: P2017P1
; CURRENT APPLICATION NUMBER: US/10/799,747
; CURRENT FILING DATE: 2004-03-15
; PRIOR APPLICATION NUMBER: US/10/195,730
; PRIOR FILING DATE: 2002-07-16
; PRIOR APPLICATION NUMBER: US/09/281,976
; PRIOR FILING DATE: 1999-03-31
; PRIOR APPLICATION NUMBER: 60/060,837
; PRIOR FILING DATE: 1997-10-02
; PRIOR APPLICATION NUMBER: 60/060,862
; NUMBER OF SEQ ID NOS: 390
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 317
; LENGTH: 149
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (128)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (133)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (136)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

FEATURE:
; NAME/KEY: SITE
; LOCATION: (140)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (143)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-799-747-317

Query Match 7.7%; Score 662; DB 16; Length 149;
Best Local Similarity 96.1%; Pred. No. 2e-29;
Matches 122; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
Qy 1250 HLSTLDRSVWKSILNARCKICRKKGDAENNVLCDCGDRGHHTYCYVRPKLKTYPGDDWF 1309
Db 12 HLSTLDRSVWKSILNARCKICRKKGDAENNVLCDCGDRGHHTYCYVRPKLKTYPGDDWF 71
Qy 1310 CPECRPQRCRRLSFRORPSLESDVDSMGDEDDVDGDEERGQSEEEYEVEQDEDD 1369
Db 72 CPECRPQRCRRLSFRORPSLESDVDSMGDEDDVDGDEERGQSEEEYEVEQDEDD 131
Qy 1370 SQEEEV 1376
Db 132 SXEEV 138

RESULT 36
US-10-094-466-30
; Sequence 30, Application US/10094466
; Publication No. US20030203363A1
; GENERAL INFORMATION:
; APPLICANT: Spytek et al.
; TITLE OF INVENTION: NOVEL HUMAN PROTEINS, POLYNUCLEOTIDES ENCODING THEM
; TITLE OF INVENTION: AND METHODS OF USING
; TITLE OF INVENTION: THE SAME
; FILE REFERENCE: 21402-2900
; CURRENT FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: US/10/094,466
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/274,281
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/288,148
; PRIOR FILING DATE: 2001-05-02
; PRIOR APPLICATION NUMBER: 60/274,849
; PRIOR FILING DATE: 2001-03-09
; PRIOR APPLICATION NUMBER: 60/275,235
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: 60/338,375
; PRIOR FILING DATE: 2001-12-04
; PRIOR APPLICATION NUMBER: 60/275,579
; PRIOR FILING DATE: 2001-03-13
; PRIOR APPLICATION NUMBER: 60/335,302
; PRIOR FILING DATE: 2001-10-31
; PRIOR APPLICATION NUMBER: 60/275,601
; PRIOR FILING DATE: 2001-03-13
; PRIOR APPLICATION NUMBER: 60/276,000
; PRIOR FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: 60/277,338
; PRIOR FILING DATE: 2001-03-20
; Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: PatIn 2.1
; SEQ ID NO 30
; LENGTH: 1398
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-094-466-30

Query Match 4.2%; Score 361.5; DB 15; Length 1398;
Best Local Similarity 18.8%; Pred. No. 2e-11;
Matches 265; Conservative 184; Mismatches 457; Indels 505; Gaps 49;
Qy 436 SQEDNVANKQTLASYSRKATKRDKLLKQEMKSLAFKAKLKREK-----ADAL 485

Db 175 SQDNSSRES--PSLEDEETKKEETPKQBEQ-----ESBKMKEEQPMDLENSTANVL 228
Qy 486 E--AKKKEKEDK-----KREELKKIVVEERLKKKEKERLKVE-----523
Db 229 BETTVKKEDEKELVKLPVVKLEKPLPENEKKIIKEESDSPKENVKPIKVVKECRA 288
Qy 524 -----REKER-----EKLREKRKYVEYLKOWSKP 548
Db 289 DPKDTKSSMEKPVNAQEPERIEFGNIVKSHSHEITEKSTEEKLNKDOQAQKIPKKREKL 348
Qy 549 REDME-----CDLDELPEPTPVKTRLPPEIFGDALMVLEFLNAFCELPDLODEPDG 601
Db 349 SDDFDSVPKGLCKSV-----TPTKEFLDKIQAQEB-----ETCKRISTITALCHEGKQL 398
Qy 602 VTLEVLBEALVGNDSGELCELLPFFLTATFOAIEEVEEVAKEQLDADTKGCSLSKLD 661
Db 399 VNGEVSDEVPAPNFKTEPI-ETKPY-----ETKEESYSPSKORNIITEGNGTESLN 448
Qy 662 --LDSCTLSLIR-LHILASGADVTSANAKRYOKRGGFDDACMELRLSNPVLVKL 718
Db 449 SVTSMKTGELEKETAPLRKAD--SSISVLIEHSQKAQIEPDPPEMETSJDSSEMAKDL 507
Qy 719 SSTSVYDLTPCEKMKILHALCGKLLTLVSTRDP---IEDYVDILRQAQEFRELKASQHR 775
Db 508 SSKTALSSTESCTMK-----GEEKSPKTKOKRPPILECLEKLEKSKTKTFLDKDAQRLS 561
Qy 776 KEREEAAARIRKKEKLEKEQKMKQ-----EKL---KEBQRNSTADISIGE 823
Db 562 PIPEE-----VPKSTLESEKFGSPAAETSPPSNIIDHCEKLAKEKVEVVEQSTSTVGQS 617
Qy 824 EEREDFTSTESKTEOKELD--QDMFTEDEDDPGSHKRRGRKRGONGFKEFTRQEQIN 881
Db 618 VKVDLETLKEDSEFTKVENMDNLNADTSGIEBSETK---GSMQSKFK-----YK 666
Qy 882 CVTRELLTADDEEALKQEHQKELKEKIQSAIACNTNIFPLGRDRMYRYWIYPPSIPGL 941
Db 667 LVPEEETTASENTITSERQKEGKILIRISS-----RKKK-----PDSPPK 708
Qy 942 FIEDYSGLTEDMLLPSPSFQNNVQSDPQVSTKGTCEPLMSESTSIDQPRHVSQLP 1001
Db 709 VLE-----PENKQKTEKEEEK--TNVGRTLR-----RSRIS 739
Qy 1002 KPVHKPNRWCYSSCEQLDOLIEALNSRGHRESALKETLLQEKSRICAQLARFSEKPHF 1061
Db 740 RPTAK-----VAEIRDQKADKRGEGEVEEESTALQK-----TDKKEI 779
Qy 1062 SDKPQDPKPTYSGRSSNAYFPSCMAEKQLELRDLFLDIEDRIYQGTGAIKVTRD 1121
Db 780 LKXKEDTNSKVK-----DEPCKKCGLPNHPILLCDSCDGYHTACLAPPL 882
Qy 1122 HWRSALESGRYELLS--EENKENGIIKTVNEDVEEMIDEQTKVIVKDRLLGKITETPST 1180
Db 800 VRMTGSRTRGRWKYSSNDESGSEKSAASEEKESEEAELADD-----847
Qy 1181 VSTNASTPQSVSSVVHYLAMALFOIEQIERFLKAPLDASDSGRSYKTVLDRWRESILS 1240
Db 848 -----847
Qy 1241 SASLSQVFLHSLTDRSVWKSILNARCKICRKKGDAENNVLCDCGDRGHHTYCYVRPKL 1300
Db 848 -----DEPCKKCGLPNHPILLCDSCDGYHTACLAPPL 882
Qy 1301 KTVPEGDWFCPECRPKORCRL-----SFRQPSLESD-----VEDSMGGE 1343
Db 883 MIIPDGEWFCPPCQHKLCEKLEQLDLVALKKERARERKRLVYVVISIENIIPPQ 942
Qy 1344 DDEVGDGEERGQSEEEYEV-----SODEDDSOE-----1372
Db 943 EPDFSEDEQEEKKSKSKANLLERRSTRTRKCIYSYRPFDEFAIDAIEDDKEADGGG 1002
Qy 1373 -----EEVSLPKRGRPQVRLPVKTRGKLSSESF-----1401

```
Db 1003 VGRGKDIITITCHRGKDIITILDEERKENKRPQAAAAARRKRRRLNLDSDSNLDEES 1062
Qy 1402 -----SRGQOQ-----PGYPSRSQOSTPKTTVSSK--TGRSLRKINSAPPTETKSLRIA 1450
Db 1063 EDEFKISDGSQDEFVVSNDENPDESEDPNDSDSDTDFCSRRLRRHSPRMRQSRRLARK 1122
Qy 1451 SRSTRHSHGPIQ-----ADVVELLSPRKRGKRSANNTPENSFP 1492
Db 1123 TPKKYSDDDDEEESSEENRSDSEDFDDDFDVE-TRRRRSRNRQKQINRYKEDSESD 1181
Qy 1493 PNFRVIATKSEQSRSVNIASK--LSLOES-----ESKRRCRKR----- 1529
Db 1182 G-----SOKSLRRGKEIRRVHVKRLLSSSEESYLSKNSEDELAKEKRSVRKRGSTD 1236
Qy 1530 ---QSPEPSVTGLRRSSRGQGVHLSAFQVLVVELVRHDDSWPFLKLVSKIQVPDYD 1586
Db 1237 EYSEADEEBEEGKPSRKR--LHRIETDEESCDNAHGDAQPARDSPRV-LPSSEQ 1292
Qy 1587 IIKPIALNIIREKNVKECYKLASEFIDIE 1617
Db 1293 STKXP-----YRIESDEEDFE 1309

RESULT 37
US-09-764-176-7
; Sequence 7, Application US/09764176
; Patent No. US20020127553A1
; GENERAL INFORMATION:
; APPLICANT: NOTEBORN, Mathieu Hubertus Maria
; APPLICANT: DANEN-VAN OORSCHOT, Astrid Adriana Anna Maria
; APPLICANT: ROHN, Jennifer Leigh
; APPLICANT: WEISS, Bertram
; TITLE OF INVENTION: APOPTIN-ASSOCIATING PROTEIN
; FILE REFERENCE: 4725US
; CURRENT APPLICATION NUMBER: US/09/764,176
; CURRENT FILING DATE: 2001-01-17
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 7
; LENGTH: 1400
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Description of Sequence: Amino acid sequence deduced from the nucleotide sequence of the cDNA clone pR1.1.
; OTHER INFORMATION: leuc acid sequence of AAP-
US-09-764-176-7

Query Match 4.2%; Score 361.5; DB 9; Length 1400;
Best Local Similarity 18.8%; Pred. No. 2e-11;
Matches 265; Conservative 184; Mismatches 457; Indels 505; Gaps 49;

Qy 436 SOEDNVANKQTILASVRSKATKRDKLLKQSEWKSIAFEKAKLKREK-----ADAL 485
Db 177 SQDNSSRES--PSLEDEBTKEETPKQESQK-----ESKKEGEEQPMLENSTANVL 230
Qy 486 E--AKKKEKEDKEK-----KRSELKVIIEERLKKKEERLKV----- 523
Db 231 EETTVKKEDEKELVKLPVIVKLEKPLPENEKKIIEESDSKPVNKVPIKVEVKCRA 290
Qy 524 -----REKER-----EKUREKKRYVEYLKQWKP 548
Db 291 DPKDTKSMKPEVAQEPERIEFGNGKSHSBEITEKSTBEETKLNDDQQAQKPLKKEIKL 350
Qy 549 REDME-----CDDLKELPPTPKTLPPEIFGDALMWLEFLNAGFELDLODEPDG 601
Db 351 SDDFDSPVKGLKSV-----TPTEKFLKDEIKQEE-----ETCKRISTITIALGHEGKQL 400
Qy 602 VTLEYLEALVGNDSQGLPCELLFFFLTAIQAIAEEREEVAKEQLTDADTKGCSLSLD 661
Db 401 VNGEVSDEKVAPEKTEPI-ETKYV-----ETKEESYSPSKORNIITEGNGTESLN 450
Qy 662 --LDSCTLSEILR-LHLILASGADVTSANAKYRQKRGGFDDTDACMELRLSNPVLVKL 718
```

```
Db 451 SVITSMKTGELEKETAPLRKAD--SSISVLIIHQAKQIEPDPPEMETSLLDSEMAKDL 509
Qy 719 SSTSVYDLTPGEKMKILHALCGKLLTVLSTRDF---IEDYVDILRQAQKQEFRELAQOHR 775
Db 510 SSKTALSTESCTWK-----GEEKSPKTKDKPPILLECLEKLEKSKTKTFLDKDAQRLS 563
Qy 776 KEREEAARIRKREKELKEQEQKMKQ-----EKL---KEDQRNSTADISIGE 823
Db 564 PIPEE---VPKSTLESEKPGSPAETSPPSIIDHCEKLASEKVEVCOSTSTVGGQS 619
Qy 824 EEREDFTSIESKDEQKELD--ODMFTEDDDPGSHKRGRRGKRGONGEFTTREQIN 881
Db 620 VKKVDLETLEKSDSFTKVEMDNLDNAOTSGIEESETK-----GSMQSKFK-----YK 668
Qy 882 CVTRELITADEEALKEHQKKEKLEKIQSAIACTNIFPLGRDRMYRYWIPFIPGL 941
Db 669 LVPEEETASENTEITSERQKEGIKLITRISS-----RKKK-----PDSPPK 710
Qy 942 FIEEDYSGLTEDMLLPRPSSPQNNVQSDPOVSTKTGEPLMSESTSNIDQGRDHVSQLP 1001
Db 711 VLE-----PENKQEKTEEEK--TNVGRTLR-----RSPRIS 741
Qy 1002 KPVHKPNRWCFYSSCEQLDOLIEALNSRGHRESALKETLLOEKSRICQAARFSEKPHF 1061
Db 742 RPTAK-----VAEIRDQKADKKRGEDEVEEESTALOK-----TDKKEI 781
Qy 1062 SDKQPDSPKPTYSRGRSSNAYDPSQCAEQLELRDLFLDIEDRIYQGTGLGAIKVTDR 1121
Db 782 LKSEKOTNSKVK-----VKPGK 801
Qy 1122 HIWRSALSGRYELLS--EENKENGIIKTVNEDVEMEIDEQTKVIVKRLLGITETPST 1180
Db 802 VRMTGSRTRGRWKYSSNDESESGSEKSSAASEEKESEEAIIADD----- 849
Qy 1181 VSTNASTPQSVSVVHYLAMALFOIEQIERRFLKAPLDASDSGRSYKTVLDRWRRESLLS 1240
Db 850 ----- 849
Qy 1241 SASLSQVFLHLSTLDRSVIWSKSIINARCKIKRKGDAENMVLCDCGRGHHTYCVRPKL 1300
Db 850 -----DFECKCGLPNHPELILLCDSCDSCGYHTACLRLPPL 884
Qy 1301 KTVPEGWDFCECPKQRCRRL-----SPQRPSLEDED-----VEDSMGE 1343
Db 885 MIIDGWFPCPPCQHKLLCEKLEQLDLVALKKERAERRKERLVVVGISIIENIIPPQ 944
Qy 1344 DDEVGDDEEGQSEEEYEV-----EQEDDSQE----- 1372
Db 945 EPDFSEDEQEEKKDSKSKANLLERRSTRTRKTCISYRPFDEPDEAIDEAIEDDIKEADGG 1004
Qy 1373 -----EEVSLPKRGRPQVRLPVKTRGKLSSEFS----- 1401
Db 1005 VGRGKDIITITCHRGKDIITILDEERKENKRPQAAAAARRKRRRLNLDSDSNLDEES 1064
Qy 1402 -----SRGQOQ---FCYPSRSQOSTPKTTVSSK--TGRSLRKINSAPPTETKSLRIA 1450
Db 1065 EDEFKISDGSQDEFVVSNDENPDESEDPNDSDSDTDFCSRRLRRHSPRMRQSRRLARK 1124
Qy 1451 SRSTRHSHGPIQ-----ADVVELLSPRKRGKRSANNTPENSFP 1492
Db 1125 TPKKYSDDDDEEESSEENRSDSEDFDDDFDVE-TRRRRSRNRQKQINRYKEDSESD 1183
Qy 1493 PNFRVIATKSEQSRSVNIASK--LSLOES-----ESKRRCRKR----- 1529
Db 1184 G-----SOKSLRRGKEIRRVHVKRLLSSSEESYLSKNSEDELAKEKRSVRKRGSTD 1238
Qy 1530 ---QSPEPSVTGLRRSSRGQGVHLSAFQVLVVELVRHDDSWPFLKLVSKIQVPDYD 1586
Db 1239 EYSEADEEBEEGKPSRKR--LHRIETDEESCDNAHGDAQPARDSPRV-LPSSEQ 1294
Qy 1587 IIKPIALNIIREKNVKECYKLASEFIDIE 1617
```

Db 1295 STKP-----YRIESDEEDFE 1311

RESULT 38

US-10-437-963-133671

; Sequence 133671, Application US/10437963

; Publication No. US20040123343A1

; GENERAL INFORMATION:

; APPLICANT: La Rosa, Thomas J.

; APPLICANT: Kowalic, David K.

; APPLICANT: Zhou, Yihua

; APPLICANT: Cao, Yongwei

; APPLICANT: Wu, Wei

; APPLICANT: Boukharov, Andrey A.

; APPLICANT: Barbazuk, Brad

; APPLICANT: Li, Ping

; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With

; FILE REFERENCE: 38-21(53221)B

; CURRENT APPLICATION NUMBER: US/10/437,963

; CURRENT FILING DATE: 2003-05-14

; NUMBER OF SEQ ID NOS: 204966

; SEQ ID NO 133671

; LENGTH: 638

; TYPE: PRT

; ORGANISM: Oryza sativa

; FEATURE:

; OTHER INFORMATION: Clone ID: PAT_MRT4530_35518C.1.pap

US-10-437-963-133671

Query Match 4.0%; Score 341; DB 16; Length 638;

Best Local Similarity 20.1%; Pred. No. 1e-10;

Matches 185; Conservative 111; Mismatches 248; Indels 376; Gaps 30;

Qy 148 GREPLLRKPFVRQPPADLRPEEVFYCKVTNEIFRHYDDFFERTILCNLSVMSCAVT 207

Db 51 GGPMLFKRPFSLLEPPKOLDSKEK-----IS 78

Qy 208 GRPLGTQEALESSEKARONLQSPPELIIIPVLYLTSLTHRSRLHEICDDIFAVVKDRYF 267

Db 79 GKSNTFEEALVSEHHAHVAKSAQKLPTELMAVPLRMIQVSTIG-LYELVEKIYASLOEAVF 137

Qy 268 VEETVEV-IRNNGARLOCTILEVLPPSHONGFANGHVNSVDGETIISDSDDSETQSCSF 326

Db 138 --EGLELYAKQDGLAEACRLIKILG-----160

Qy 327 QNGKKDAIDPLLPKYKVQPTKELHESAIVKATQISRRKHLFRDRLKLPKQHCBPQE 386

Db 161 SDGTMVEVGWLL-----RDKTIISTSVIKGEDLIHRRPPVSRNTLKIFIRD-----207

Qy 387 GVIKIKASSLSYKIAQDFDS--YFFDDPPT-FIFSPANRRGRPPKRIHISOEDNVAN 443

Db 208 -----ATSQNAPWVIHENLAKRYGPIEPENDMMFGEGLQKGRK-----247

Qy 444 KQTLASYSRKATKRDKLLKQEMKSLAFKAKREKADALEAKKKEKEDKREELK 503

Db 248 -----RRDGPMDPKKKQKKDEEHINVPK 273

Qy 504 KIVEERLKKKEKERLKEREREKRLREBKRYVEYLKQWSKPREDMECDLKEPPEP 563

Db 274 YPIDD-----LLVQ-----PSADDHALLKRP 294

Qy 564 TPVKT--RLPPEIRGDALMWLEFLNARGELFDLQDFPDGVTLEVLREALVGNDSSEGLC 621

Db 295 -PLATDFRPVKYSYVGLLMMWDFCLSEGRVNLSP-----FSLVDLENAICHKESNALLV 348

Qy 622 ELLFFFLTAIFOAAEEREEVAKQLTADATKGCGLKSLDLDSTLSEILRLHLIASSAD 681

Db 349 EI-----HTAIFHLIKOE-----GYFILRTKKRKLK-----VLTWTAEYLCDFLE 392

Qy 682 VTSANAKRYOKRGGFATDDACMELNSPLVKLSSTSYVOLTPEE-KMKILHALCG 740

Db 393 MTKTEELTR-----NIATVRK-----GYSLIDTDIKLILRELVE 428

Qy 741 KLLTLVSTRFIEDYVDILRQAKQEFFRELKAEQHKEREREAARIRKRKEKLEKEQOKM 800

Db 429 EAITTSPVREKLSERVD-----QROLAATKRESTRKAKDEQNSSI-----DGLQDDNESV 479

Qy 801 KEQEKLEDEQRNSTADISIGBEEREDFDTISBESKOTEQKELQDQDMFTDEDDPGSHKR 860

Db 480 DE-QGKGKEEKDKN-----NISRSKTEGK-----502

Qy 861 GRRGKRGQNGFKEPTROEQINCVTRELLTADBEALQEHQKKEKLEKIQSAIACTNI 920

Db 503 -RHG-----VOHLETEIEKL-SIRSS-----521

Qy 921 FPLGRDRMYRYWIFPSIPGLFIEEDYSGLTEDMLLRPSSFQNNVQSDPQVSTKTGEP 980

Db 522 -PLGKORHYNRYWFFRREGRLFVESADS-----548

Qy 981 LMSESTSNIDQGRDRHSVOLPKPVHKPNRMCFYSSCQLDQLEALNSRHRGHSALKETL 1040

Db 549 -----KNGYYSTKEELDVLMSSLVNKGRLRERALKRQL 581

Qy 1041 LOEKSRICAOIARFSEKFKH 1060

Db 582 DKLYSKISNALEKRSKEITH 601

RESULT 39

US-09-864-761-34546

; Sequence 34546, Application US/09864761

; Patent No. US20020048763A1

; GENERAL INFORMATION:

; APPLICANT: Penn, Sharon G.

; APPLICANT: Rank, David R.

; APPLICANT: Hanzel, David K.

; APPLICANT: Chen, Wensheng

; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR

; FILE REFERENCE: Acomica-X-1

; CURRENT APPLICATION NUMBER: US/09/864,761

; CURRENT FILING DATE: 2001-05-23

; PRIOR APPLICATION NUMBER: US 60/180,312

; PRIOR FILING DATE: 2000-02-04

; PRIOR APPLICATION NUMBER: US 60/207,456

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: US 09/632,366

; PRIOR FILING DATE: 2000-08-03

; PRIOR APPLICATION NUMBER: GB 24263.6

; PRIOR FILING DATE: 2000-10-04

; PRIOR APPLICATION NUMBER: US 60/236,359

; PRIOR FILING DATE: 2000-09-27

; PRIOR APPLICATION NUMBER: PCT/US01/006666

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/006667

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/006664

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/006669

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/006665

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/006668

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/006663

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/006662

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/006661

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/006670

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: US 60/234,687

; PRIOR FILING DATE: 2000-09-21

; PRIOR APPLICATION NUMBER: US 09/608,408

;; PRIOR FILING DATE: 2000-06-30
;; PRIOR APPLICATION NUMBER: US 09/774,203
;; PRIOR FILING DATE: 2001-01-29
;; NUMBER OF SEQ ID NOS: 49117
;; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
;; SEQ ID NO 34546
;; LENGTH: 572
;; TYPE: PRT
;; ORGANISM: Homo sapiens
;; FEATURE:
;; OTHER INFORMATION: MAP TO AC005089.2
;; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3
;; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 5.4
;; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 4
;; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 5.6
;; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.3
;; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.1
;; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 4
;; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.7
;; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3.1
;; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 3.7
;; OTHER INFORMATION: SWISSPROT HIT: P53935, EVALUE 2.00e-03
;; OTHER INFORMATION: EST_HUMAN HIT: AW976211.1, EVALUE 2.00e-91
US-09-864-761-34546

Query Match 3.7%; Score 318.5; DB 9; Length 572;
Best Local Similarity 24.8%; Pred. No. 1.6e-09;
Matches 124; Conservative 89; Mismatches 173; Indels 115; Gaps 20;

Qy 429 PPKRIHI-----SQEDNVANKOTLASYSRKATKDKLLKOEEMKSLAFKAKLKREKAD 483
Db 175 PPAALHLIAYKENDKRDKSALSCVSKTA-----RLLSSEDRARLPEELRSLVQKYE 230
Qy 484 ALEAKK-EKEDKEKKREELKKIVEERLKKKEERLKKVEREKREKLREKRYEYL 542
Db 231 LLEHKRWASNEEQKEYLKK--KREELKKK-LKEKAKERKEKEMLEKLEKQRY-----283
Qy 543 KOWSKPRMECDUUKELPEPTPVKT--RLPPEIPGDALMVLFNAGELFDLODEPPD 600
Db 284 -----BDQELTG-KNLPAFLVDTPGCLPNTLFGDVAMVVEFLSCYSGLLLPDAQYP- 334
Qy 601 GTVLVLEALVGNDSQPLCELLPFFLTAFQAIABE--EEVAKELQTDADTKGCSLK 658
Db 335 -ITAVSLMEAL-SADKGG-----FLYNRLVILLQTLQDEIAB-----DYGELGMKLS 382
Qy 659 SLDDSCSLTBILRHILASGADVTSANAKRYQKRGFGFATDDACMELRLSNPSLVKLL 718
Db 383 EIPLTHSVSELVRLCLRSQVSESGSDTD-----DNKDSAAFEDNEVQDEFLEKL 435
Qy 719 SSTSYDLTPGEKMKILHALCGKLLTLVSTRDFIEDYVDVILRQAKQEFRELKASQHRKER 778
Db 436 ETSBFFELTSBEKLQILTALCHRLMTYSVD-----HMETR 472
Qy 779 EEAARIRKREKLEKEQKQKQKQKQKLEDEQNRSTADISIGEEREDFTSIESKDT 838
Db 473 QQMSAELKRLAVLKEKDKKRAKQKQK-----MEAKNK 509
Qy 839 EOKELQDMFTEDDDPSHKRGRGKRGQNGFKFTROE-QINCVTRELLTADDEEALK 897
Db 510 ENGVEN-----GLGTDRK-----KEIVKFFQVDTAEADMISAVKSRLL 551
Qy 898 QEHQRKEKELLEK-IQSAIAC 917
Db 552 AIQAKKEREIQEREMKGKISC 572

RESULT 40
US-09-864-761-37938
; Sequence 37938, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.

;; APPLICANT: Hanzel, David K.
;; APPLICANT: Chen, Wensheng
;; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
;; FILE REFERENCE: Aecomica-X-1
;; CURRENT APPLICATION NUMBER: US/09/864,761
;; CURRENT FILING DATE: 2001-05-23
;; PRIOR APPLICATION NUMBER: US 60/180,312
;; PRIOR FILING DATE: 2000-02-04
;; PRIOR APPLICATION NUMBER: US 60/207,456
;; PRIOR FILING DATE: 2000-05-26
;; PRIOR APPLICATION NUMBER: US 09/632,366
;; PRIOR FILING DATE: 2000-08-03
;; PRIOR APPLICATION NUMBER: GB 24263.6
;; PRIOR FILING DATE: 2000-10-04
;; PRIOR APPLICATION NUMBER: US 60/236,359
;; PRIOR FILING DATE: 2000-09-27
;; PRIOR APPLICATION NUMBER: PCT/US01/00666
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00667
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00664
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00669
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00665
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00668
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00663
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00662
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00661
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00670
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: US 60/234,687
;; PRIOR FILING DATE: 2000-09-21
;; PRIOR APPLICATION NUMBER: US 09/608,408
;; PRIOR FILING DATE: 2000-06-30
;; PRIOR APPLICATION NUMBER: US 09/774,203
;; PRIOR FILING DATE: 2001-01-29
;; NUMBER OF SEQ ID NOS: 49117
;; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
;; SEQ ID NO 37938
;; LENGTH: 560
;; TYPE: PRT
;; ORGANISM: Homo sapiens
;; FEATURE:
;; OTHER INFORMATION: MAP TO AC005074.1
;; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.2
;; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.5
;; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2.8
;; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.1
;; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.9
;; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.93
;; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.3
;; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.8
;; OTHER INFORMATION: EST_HUMAN HIT: AW976211.1, EVALUE 2.00e-91
;; OTHER INFORMATION: SWISSPROT HIT: P53935, EVALUE 2.00e-03
US-09-864-761-37938

Query Match 3.6%; Score 314; DB 9; Length 560;
Best Local Similarity 24.7%; Pred. No. 2.8e-09;
Matches 122; Conservative 86; Mismatches 171; Indels 114; Gaps 19;
Qy 429 PPKRIHI-----SQEDNVANKOTLASYSRKATKDKLLKOEEMKSLAFKAKLKREKAD 483
Db 168 PPAALHLIAYKENDKRDKSALSCVSKTA-----RLLSSEDRARLPEELRSLVQKYE 223
Qy 484 ALEAKK-EKEDKEKKREELKKIVEERLKKKEERLKKVEREKREKLREKRYEYL 542

Db 224 LLEHKRWASNSEQRKEYLKK--KREELKKK-LKEKAKERREKEMLERLEKQKRY---- 276
Qy 543 KQWSPREDMECDLDELPEPTPVKT--RLPEIFGDALAVLEFNLARGSELFDLQDEPPD 600
Db 277 -----BDQBLTG-KNLPFAFLVDTPRGLPNTLFGDVAMTVFELSCYSGULLPDAQYP- 327
Qy 601 GVTEVLREALVGNDSQPLCELLFFFLTAIFQAIABE--EEVAKGOLTDADTKGCSLK 658
Db 328 -ITAVSLMEAL-SADKGG-----FLYNRVILVILLQTLQDEIAE----DYSELGWMKLS 375
Qy 659 SLDDSCSLSELRLHLILASGADVTSANAKYRYOKRGGFDDTADACMELRLSNLSNVLKVL 718
Db 376 BIPLTLHSVSELVRLCLRRSDVQSESGSDT-----DNKDSAAFEDNEVQDEFLEKL 428
Qy 719 SSTSVYDLTPGKWKILHALCGKLLTVSTDRDFTEDYVDILRQAKQEFRELKAKQHRKR 778
Db 429 ETSEFFELTSBEKLQILTALCHRLMTYSVQD-----KEIVKFEPOVTEADDMISAVKSRLL 465
Qy 779 EEAARIRKREKLEKEQKWKKEQKLEKEDEQNSTADISIGEEREDFDTSIESKOT 838
Db 466 QOMGAELWKERLAVLKEENDKKRAEKQKRKE-----MEAKNK 502
Qy 839 EQKELDDMFTEDEDDPGSHKRGRRGKRGQNGKFEFTQOE-QINCVTRELLTADDEEALK 897
Db 503 ENGVEN-----GLGKTRK-----KEIVKFEPOVTEADDMISAVKSRLL 544
Qy 898 QEHQKKEKELLEK 910
Db 545 AIQAKEREIQER 557

RESULT 41
US-09-839-479-37
; Sequence 37, Application US/09839479
; Publication No. US20020039779A1
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042002
; CURRENT APPLICATION NUMBER: US/09/839,479
; CURRENT FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 37
; LENGTH: 59
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-839-479-37

Query Match 3.6%; Score 313; DB 9; Length 59;
Best Local Similarity 100.0%; Pred. No. 1.9e-10;
Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1569 SWPFLKLVSKIQVDPDYDIKKPIALNIIREKVNKCEYKLASEFIDDIELMFNSCFEYN 1627
Db 1 SWPFLKLVSKIQVDPDYDIKKPIALNIIREKVNKCEYKLASEFIDDIELMFNSCFEYN 59

RESULT 42
US-09-839-479-49
; Sequence 49, Application US/09839479
; Publication No. US20020039779A1
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042001
; CURRENT APPLICATION NUMBER: US/10/376,537

; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042002
; CURRENT APPLICATION NUMBER: US/09/839,479
; CURRENT FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 49
; LENGTH: 59
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-839-479-49
Query Match 3.6%; Score 313; DB 9; Length 59;
Best Local Similarity 100.0%; Pred. No. 1.9e-10;
Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1569 SWPFLKLVSKIQVDPDYDIKKPIALNIIREKVNKCEYKLASEFIDDIELMFNSCFEYN 1627
Db 1 SWPFLKLVSKIQVDPDYDIKKPIALNIIREKVNKCEYKLASEFIDDIELMFNSCFEYN 59

RESULT 43
US-10-376-537-38
; Sequence 38, Application US/10376537
; Publication No. US20030224405A1
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042001
; CURRENT APPLICATION NUMBER: US/10/376,537
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US/09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 38
; LENGTH: 59
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-376-537-38

Query Match 3.6%; Score 313; DB 15; Length 59;
Best Local Similarity 100.0%; Pred. No. 1.9e-10;
Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1569 SWPFLKLVSKIQVDPDYDIKKPIALNIIREKVNKCEYKLASEFIDDIELMFNSCFEYN 1627
Db 1 SWPFLKLVSKIQVDPDYDIKKPIALNIIREKVNKCEYKLASEFIDDIELMFNSCFEYN 59

RESULT 44
US-10-376-537-50
; Sequence 50, Application US/10376537
; Publication No. US20030224405A1
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042001
; CURRENT APPLICATION NUMBER: US/10/376,537

```
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US/09/418,710
; PRIOR FILING DATE: 1998-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 50
; LENGTH: 59
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-376-537-50

Query Match          3.6%; Score 313; DB 15; Length 59;
Best Local Similarity 100.0%; Pred. No. 1.9e-10;
Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1569 SWPFLKLVSKIQVPDYDIKKPIALNIIRKVNKCEYKLASEFIDDIELMFNSCFEYN 1627
Db 1 SWPFLKLVSKIQVPDYDIKKPIALNIIRKVNKCEYKLASEFIDDIELMFNSCFEYN 59

RESULT 45
US-10-702-148-37
; Sequence 37, Application US/10702148
; Publication No. US20040063145A1
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042002
; CURRENT APPLICATION NUMBER: US/10/702,148
; CURRENT FILING DATE: 2003-11-05
; PRIOR APPLICATION NUMBER: PRIOR APPLICATION NUMBER: US/09/839,479
; PRIOR FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 37
; LENGTH: 59
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-702-148-37

Query Match          3.6%; Score 313; DB 15; Length 59;
Best Local Similarity 100.0%; Pred. No. 1.9e-10;
Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1569 SWPFLKLVSKIQVPDYDIKKPIALNIIRKVNKCEYKLASEFIDDIELMFNSCFEYN 1627
Db 1 SWPFLKLVSKIQVPDYDIKKPIALNIIRKVNKCEYKLASEFIDDIELMFNSCFEYN 59

RESULT 46
US-10-702-148-49
; Sequence 49, Application US/10702148
; Publication No. US20040063145A1
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042002
; CURRENT APPLICATION NUMBER: US/10/702,148
; CURRENT FILING DATE: 2003-11-05
```

```
; PRIOR APPLICATION NUMBER: PRIOR APPLICATION NUMBER: US/09/839,479
; PRIOR FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 49
; LENGTH: 59
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-702-148-49

Query Match          3.6%; Score 313; DB 15; Length 59;
Best Local Similarity 100.0%; Pred. No. 1.9e-10;
Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1569 SWPFLKLVSKIQVPDYDIKKPIALNIIRKVNKCEYKLASEFIDDIELMFNSCFEYN 1627
Db 1 SWPFLKLVSKIQVPDYDIKKPIALNIIRKVNKCEYKLASEFIDDIELMFNSCFEYN 59

RESULT 47
US-10-335-977-7982
; Sequence 7982, Application US/10335977
; Publication No. US20040052799A1
; GENERAL INFORMATION:
; APPLICANT: DOUGLAS SMITH et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
; RELATING TO HELICOBACTER PYLORI FOR
; DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 10031
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: Windows NT 4.0
; SOFTWARE: UNIX
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/335,977
; FILING DATE: 30-Dec-2002
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/993,002
; FILING DATE: 17-DEC-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragouras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: GTN-018
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)742-4214
; INFORMATION FOR SEQ ID NO: 7982:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1820 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: misc_feature
```

; LOCATION: (B) LOCATION 1...1820	
; SEQUENCE DESCRIPTION: SEQ ID NO: 7982:	
US-10-335-977-7982	
Query Match	
Beat Local Similarity 3.6%; Score 310.5; DB 15; Length 1820;	
Matches 278; Conservative 225; Mismatches 511; Indels 423; Gaps 63;	
Qy	312 IISDSDSETQSCFQNGKKDAADPL-----LFGYKQPTKKELHESAIV 357
Db	213 ITDSDNDEI-----IKSGKKYIIGGIVAVLIVILFRSIFHYFV-PLSD----- 259
Qy	358 KATQISRRKHLFSRDKLKL-----FLKQHCEPQGVKIKKASSLSTYKIAEODFSVPFP 411
Db	260 KSSRFKDRNLVYNDQIROYENRLLKERNE-----KGNMI-----DKNLFN 303
Qy	412 DDPTTFIPSPAN-----RRRGPRPKRIHISOEDNVAN-----KOTLASV-- 450
Db	304 DDPNRTLNYLNIAIEDKNPLRAFYECISNGGNVEECLKLIKOKKLQDQMKKTLEAYND 363
Qy	451 ---RSKATKERDK---LLKQEMKSLAPEKAKLREKAD-ALBAKKKEKEDKREBELK 503
Db	364 CIKNAKTEERIKCLDLIKDENLK-----KSLAQKQVQVALDCLKNAKTDEERK--ECL 416
Qy	504 KIVEERLKKKEKE-----RLKVEREK-----EREKLEERKKVVEY 541
Db	417 KLINDPEIRKRELELELOKLEQYKDCIKNAKTEAKNECKLKGLSKEAIELKQQALDC 476
Qy	542 LKQMSKPREDMECDLKLPEPTPVKTRLPPEIFGDALMWLEFLNAGELFDLQDFPDG 601
Db	477 LKNAKTDEEREC--LKNIPQ-----DLQKELLAD 504
Qy	602 VTLEVEAL--VGNDSGPICELLFFFLTALFOAIAEBEE-EVAKQLTDADTKGSLK 658
Db	505 MSVKAYKDCVRARNEKEKQCEKL---LTPPEAKKLLLENQALDCLKNAKTDEEREC-LK 560
Qy	659 SLDLDSCTLSILRLHILASGADVTSANAKVRYOKRGGFDDTADACMELRSLNPSLVKL 718
Db	561 NLPKD--LQSDILAKESLAKYDCAS-QAKTEAEK-----ECKL----- 598
Qy	719 SSTSYDILTPGCKMILHALCGKLLTLVSTRDFIEDVYDILRQAKQEPRELKAE-----Q 773
Db	599 -----LTP-EAKLLEE-----EAKESVKAYLDCVSAQTEAEKCEKLLTPE 641
Qy	774 HRKREEREAARIR-----KRKEELKEQEQKKEQKLEKEDQBNSTADISIGEE 825
Db	642 AKKLEEAKKSVRAYLDCVSAKNAEAEKCEKLLTPEAKKLLLENQALDCLKNAKTDEER 701
Qy	826 RE-----DFDTSIESKDTQKELDQDMFTEDDD-----PGSHKRGRRGKRGON 869
Db	702 KECLKDLPKDLQKVLAKESVRVYLDVCSKAKNAEAEKCEKLLTPEARKLLEAAKSVK 761
Qy	870 GFKEFTREQINCVR-----ELLTADDEEAL-----KQEHQ 901
Db	762 AYKD-----CVSRARNEKEKQCEKLLTPEARKLLEESKSVKAYLDCVSAKNAEAE 813
Qy	902 RKE-----KELLEKIQSAIA-----CTNIFPLGRDMYRRTWIIPSPGLFIEDYS 948
Db	814 RKECEKLLTPEARKLLEAAKESVKAYKDCVSRARNEKEKQCEKLLTPEARKLLEESKKS 873
Qy	949 GLTEDMLLPSPSSFONNVQSDQPVSTKTGTPLMASESTSNIDQGRPHSVQLPKPVHKPN 1008
Db	874 -----VKAYLDCVSAKNAEAEKCEKLLTPEARKLLEAAKESVKAYKDCVSRAR 923
Qy	1009 RWCYSSCEQL-----DQLEAL-NSRGHRE-----SALKETLLQEKSR 1047
Db	924 NEKEKQCEKLLTPEAKKLLLENQALDCLKNAKTEAEKKRCVKDLPKDLQKVLAKESVRV 983
Qy	1048 ---CAQLARFSEKPFHFDKPOPSKPTYSGRGS-NAYDPSQWCAEQLELRDLRDLFD 1103
Db	984 YLDCVSAKNAEAEKCEKLLTPEARKLLEAAKESVKAYKDCVSRARNEKEKQCEKLLT 1043
Qy	1104 IEDR--IYQGTGAIKVTDRIHWSALESGRYE---LLSEENK---ENGII-----KTV 1149

Db	1044 PEARKLEQEVKSVKAYLDCVSRARNEKEKQCEKLLTPEARKLLENOALDCLKNAKTE 1103
Qy	1150 NED---VEEMEIDBOTKVIIVKDRLLGI-----KTTPTSTVSTNASTPQ 1189
Db	1104 AEKRCVKDLPKDLQKVLAKESVKAYLDCVSRARNEKEKKECEKLLTPEARKLLEBSKK 1163
Qy	1190 SVSSVHYLAWALFOIQGIERRFL-----KAPLDASDSGRSYKTVLDRWR-----E 1236
Db	1164 SVRAYLDCVSAKNAEAEKCEKLLTPEARKLLEAAKESVKAYKDCVSRARNEKEKQCE 1223
Qy	1237 SLLSSASLOVFLHLSTRSVIWSKSIILNARCKICRKGDAENMVLCDGCDRGHTTYCV 1296
Db	1224 KLLTPEA-----RKLEQEV--KXSVKAYLDCVSRARNEKE----- 1257
Qy	1297 RPKLUTVPEDWFPCEPKQRCRL-----SPRQPSLESDEDDVDSMGGEDEVDGD 1350
Db	1258 -----KQCEKLLTPEARKLEKQROQKQKAIKDLKNAD---PND 1295
Qy	1351 EE-----EGOSEEEYEVEQDEDDSOEEEVSLPKRGRQVRLPVKTRGKLSSESFSRG 1404
Db	1296 RAAIMKCLDGLSDEBKYLQ---EAREKAVLDCCLKTARTD-----BEKRCQNDLYSOLI 1347
Qy	1405 QQOEPRGYPSRSQOSTPKTTVSSKTRSLRKINSAPPTETKSLRIASRSTRSHSGPL--- 1461
Db	1348 QEIQ-NKKAQNKNQLSKTERLHQASECLDNLDD--PTDQEAIEQCLEGLSDSERALLIG 1404
Qy	1462 ---QADVVELLSPPRKRR--GRKSANNTPENSNNFNFRVIATKSSESQSRSVNIAS 1513
Db	1405 IKROADEVDRIYDLRSRKTFDNMAAGYPILLPMDPKNGGDIATINATNVDADKIAS 1461

RESULT 48

US-10-335-977-7981

Sequence 7981, Application US/10335977

Publication No. US20040052799A1

GENERAL INFORMATION:

APPLICANT: DOUGLAS SMITH et al

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO HELICOBACTER PYLORI FOR DIAGNOSTICS AND THERAPEUTICS

NUMBER OF SEQUENCES: 10031

CORRESPONDENCE ADDRESS:

ADDRESSEE: LAHIVE & COCKFIELD

STREET: 28 State Street

CITY: Boston

STATE: Massachusetts

COUNTRY: USA

ZIP: 02109-1875

COMPUTER READABLE FORM:

MEDIUM TYPE: CD-ROM ISO9660

COMPUTER: IBM PC Compatible

OPERATING SYSTEM: Windows NT 4.0

SOFTWARE: UNIX

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/335,977

FILING DATE: 30-Dec-2002

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/993,002

FILING DATE: 17-DEC-1997

ATTORNEY/AGENT INFORMATION:

NAME: Mandragouras, Amy E.

REGISTRATION NUMBER: 36,207

REFERENCE/DOCKET NUMBER: GTN-018

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 227-7400

TELEFAX: (617) 742-4214

INFORMATION FOR SEQ ID NO: 7981:

SEQUENCE CHARACTERISTICS:

LENGTH: 1819 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

```

; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...1819
; SEQUENCE DESCRIPTION: SEQ ID NO: 7981:
US-10-335-977-7981

Query Match      3.6%; Score 309.5; DB 15; Length 1819;
Best Local Similarity 19.4%; Pred. No. 2.2e-08;
Matches 277; Conservative 225; Mismatches 520; Indels 407; Gaps 62;

Qy 312 IISDSDSETQSCSFQNGKKKDAIDPL-----LFYKVKQPTKKELHESAIV 357
Db 212 ITDSDNDEI-----IKGSKKYYIGGVAVLVIILFSRSIFHYFV-PLD----- 258
Qy 358 KATQISRRKLFSDRLKLF-----FLKQHCPEQGVIKIKASSLSYTKIAEODFSYFFP 411
Db 259 KSRFSKDRNLVYNDEIQIQEYNRLKERNE-----XGNI-----DKNLFFN 302
Qy 412 DDPPTFIIPSPAN--RRRGRPKRIHISOEDNVAN-----KOTLASV-- 450
Db 303 DDPNRTLYNVLNIAEIEDKNPLRAFECISNGNGVYEECLIKIKDKLQDQMKKTLEAYND 362
Qy 451 ---RSKATKERDK---LLKQEMKSLAFEXAKLREKAD-ALEAKKEKEDKKEKRELK 503
Db 363 CINKNAKTEERIICLDLIKDENLK-----XSLNQQKQVQVALDCLCNKNAKTDEERK--ECL 415
Qy 504 KIVVEEBLLKKKEEKE-----RLKVEREK-----BREKLEERKKRVVEY 541
Db 416 KLINDPIREKFRKELELOKELQYKDCIKNAKTEAEKNECLKGLSKEATERLKQQALDC 475
Qy 542 LKQWSKPRDMCEDDLKELPEPTPVKTRLPPEIFGDALMVLEFUNAIFGLFDLQDEPPDG 601
Db 476 LKNAKTDEERKEC--LKNIPQ-----DLQKELLAD 503
Qy 602 VTLEVLBEAL--VGNDSGLPCELLFFFLTAIFOAIAEEBE-EVAKEQLTADTKGGSIAK 658
Db 504 MSVKAYKDCVSRARNEKEOCEKL---LTPKAKKJLENQALDCLCNKNAKTDEERKEC-LK 559
Qy 659 SLDLSDCTLSEILRLHLILASGADVTSANAKRYOKRGGFDTADACMELRLSNPSLVKKL 718
Db 560 NLPKD--LQSDILAKESLKAYKDCAS-OAKTEAEK-----ECEKL----- 597
Qy 719 SSTSVYDLTPGERKWIHALGKULLTLVSTRDFTIEDYVDILRQAKQBFRELKAE---Q 773
Db 598 -----LTP-EAKKLEEE-----EAKESVKAYLDCVSAQTEAEKKECEKLLTPE 640
Qy 774 HRKREEAARIR-----KRKEEKLKEQEKMKKEQEKLKEDDEQNSTADISIGEEE 825
Db 641 AKKLEEAASVRAYLDCVSKAKNEARKECEKLLTPEAKKJLENQALDCLCNKNAKTDEER 700
Qy 826 RE-----DFDTSIESKDTQEKELDDMFTEDEDD-----PGSHKRRGRKRGQN 869
Db 701 KECLKDPLQDLQKVLAKESVRVYLDVCVSKAKNEARKECEKLLTPEARKLLEAKKSVK 760
Qy 870 GFKEF---TROEQINCVTRELLTADBEAL-----KQEHQRKE----- 904
Db 761 AYKDCVLRARNEKEQCEKLLTPEARKLLEESKKSVKAYLDCVSKAKNEARKECEKLL 820
Qy 905 ---KELLEKIQSAIA---CTNIFPLGRDWMYRYWIFPSIPGLFTIEDYSGLTEDMLL 956
Db 821 TPEARKLLEEAESVKAYKDCVSRARNEKEQCEKLLTPEARKLLEESKKS----- 872
Qy 957 PRPSSFQNNQSQDPOVSTTKGTGPELPMSESTNIDQGRDHSVOLPKVPHKPNRMCFYS 1016
Db 873 --VKAYLDCVSKAKNEARKECEKLLTPEARKLLEEAESVKAYKDCVSRARNEKEQEC 930
Qy 1017 EQL-----DQILEAL-NSRGHRE-----SALKETLLQEKSGRI---CAQLA 1052
Db 931 EKLTLTPAKKJLENQALDCLCNKNAKTEAEKRCVKQDLPKDLQKVKLAKESVRVYLDVCVSKA 990

```


THIS PAGE BLANK (USPTO)

GenCore version 5.1.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 9, 2005, 13:54:12 ; Search time 29.454 Seconds
(without alignments)
4242.638 Million cell updates/sec

Title: US-10-702-148-1
Perfect score: 8631
Sequence: 1 MEDASESRGVAPLINNVVL.....VTPSNVDQVSTPPAAKXSRI 1674

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 65 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	8631	100.0	1674	US-09-418-710-1	Sequence 1, Appli
2	8631	100.0	1674	US-09-839-479-1	Sequence 1, Appli
3	8588.5	99.5	1673	US-09-418-710-70	Sequence 70, Appl
4	8588.5	99.5	1673	US-09-839-479-69	Sequence 69, Appl
5	1087	12.6	211	US-09-418-710-63	Sequence 63, Appl
6	1087	12.6	211	US-09-839-479-62	Sequence 62, Appl
7	954.5	11.1	1540	US-09-949-016-7037	Sequence 7037, Ap
8	930.5	10.8	1531	US-09-418-710-29	Sequence 29, Appl
9	930.5	10.8	1531	US-09-839-479-29	Sequence 29, Appl
10	929.5	10.8	1527	US-09-418-710-27	Sequence 27, Appl
11	929.5	10.8	1527	US-09-839-479-27	Sequence 27, Appl
12	926	10.7	1525	US-09-418-710-69	Sequence 69, Appl
13	926	10.7	1525	US-09-839-479-68	Sequence 68, Appl
14	726.5	8.4	1972	US-09-418-710-21	Sequence 21, Appl
15	726.5	8.4	1972	US-09-839-479-21	Sequence 21, Appl
16	704	8.2	1969	US-09-418-710-72	Sequence 72, Appl
17	704	8.2	1969	US-09-839-479-71	Sequence 71, Appl
18	682.5	7.9	1876	US-09-418-710-71	Sequence 71, Appl
19	682.5	7.9	1876	US-09-839-479-70	Sequence 70, Appl
20	671.5	7.8	1878	US-09-418-710-13	Sequence 13, Appl
21	671.5	7.8	1878	US-09-839-479-13	Sequence 13, Appl
22	361.5	4.2	1400	US-09-764-176-7	Sequence 7, Appli
23	313	3.6	59	US-09-418-710-38	Sequence 38, Appl
24	313	3.6	59	US-09-418-710-50	Sequence 50, Appl
25	313	3.6	59	US-09-839-479-37	Sequence 37, Appl
26	313	3.6	59	US-09-839-479-49	Sequence 49, Appl
27	310	3.6	1898	US-08-056-200-94	Sequence 94, Appl

28	310	3.6	1898	2	US-08-800-644-94	Sequence 94, Appl
29	310	3.6	1898	4	US-09-538-092-1280	Sequence 1280, Ap
30	309	3.6	800	4	US-09-270-767-45282	Sequence 45282, A
31	305	3.5	2107	4	US-09-949-016-7646	Sequence 7646, Ap
32	305	3.5	2107	4	US-09-949-016-7647	Sequence 7647, Ap
33	297	3.4	2101	1	US-08-466-390-4	Sequence 4, Appli
34	297	3.4	2101	1	US-08-470-950-4	Sequence 4, Appli
35	297	3.4	2101	1	US-08-467-781-4	Sequence 4, Appli
36	297	3.4	2101	1	US-08-195-487-4	Sequence 4, Appli
37	297	3.4	2101	2	US-08-483-924-4	Sequence 4, Appli
38	297	3.4	2101	3	US-09-452-294-1	Sequence 1, Appli
39	297	3.4	2101	5	PCT-US93-06160-4	Sequence 4, Appli
40	293.5	3.4	2662	4	US-09-595-6846-31	Sequence 31, Appl
41	293.5	3.4	2663	4	US-09-538-092-1252	Sequence 1252, Ap
42	288.5	3.3	2468	4	US-09-976-594-726	Sequence 726, App
43	288.5	3.3	2468	4	US-09-538-092-1135	Sequence 1135, Ap
44	288.5	3.3	2522	4	US-09-949-016-10237	Sequence 10237, A
45	281	3.3	2047	4	US-09-949-016-7404	Sequence 7404, Ap
46	279	3.2	3248	1	US-08-353-700-1	Sequence 1, Appli
47	279	3.2	3248	5	PCT-US95-16216-1	Sequence 1, Appli
48	276	3.2	45	4	US-09-418-710-43	Sequence 43, Appl
49	276	3.2	45	4	US-09-418-710-56	Sequence 56, Appl
50	276	3.2	45	4	US-09-839-479-42	Sequence 42, Appl
51	276	3.2	45	4	US-09-839-479-55	Sequence 55, Appl
52	275	3.2	1979	4	US-09-949-016-6468	Sequence 6468, Ap
53	273	3.2	1857	4	US-09-917-254-91	Sequence 91, Appl
54	273	3.2	1972	4	US-09-538-092-1084	Sequence 1084, Ap
55	273	3.2	1984	4	US-09-949-016-7111	Sequence 7111, Ap
56	273	3.2	1984	4	US-09-949-016-7112	Sequence 7112, Ap
57	273	3.2	1984	4	US-09-949-016-7113	Sequence 7113, Ap
58	269.5	3.1	3210	4	US-09-538-092-1154	Sequence 1154, Ap
59	268	3.1	1960	4	US-09-949-016-10872	Sequence 10872, A
60	267.5	3.1	2482	1	US-08-328-254-6	Sequence 6, Appli
61	267	3.1	1960	4	US-09-538-092-1077	Sequence 1077, Ap
62	267	3.1	3878	4	US-09-914-259-11	Sequence 11, Appl
63	266.5	3.1	1972	4	US-08-875-435B-3	Sequence 3, Appli
64	266	3.1	2954	4	US-09-150-867-1	Sequence 1, Appli
65	264.5	3.1	1976	4	US-09-538-092-1078	Sequence 1078, Ap

ALIGNMENTS

RESULT 1
US-09-418-710-1
; Sequence 1, Application US/09418710
; Patent No. 6596482
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042001
; CURRENT APPLICATION NUMBER: US/09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JF98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1674
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-418-710-1

Query Match 100.0%; Score 8631; DB 4; Length 1674;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1674; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MEDASESRGVAPLINNVVLPGSLPLSVVTGCKSHRVANKVAREKLLPTALPPSE 60
|||||

Db 1 MEDASESRGVAPLINNVLPGLSPLSPVSVTGCKSHRVANKVEARSEKLLPTALPPSE 60
Qy 61 PKVDOKLPRSSERRSGSGGTOPPARSRAVAAGEAARGAAGPERGSPGLGRVSPRCLCSG 120
Db 61 PKVDOKLPRSSERRSGSGGTOPPARSRAVAAGEAARGAAGPERGSPGLGRVSPRCLCSG 120
Qy 121 EGGQVAVGIAGKRRGRGDSRRAPGGREMPHLHRKPFVQKPPADLRPDVEEYFCYKT 180
Db 121 EGGQVAVGIAGKRRGRGDSRRAPGGREMPHLHRKPFVQKPPADLRPDVEEYFCYKT 180
Qy 181 NEIFRHYDDFFERTILCNLSVMSCAVTRPGGLTYQEALESKAKQNLQSPPEPLIIPVL 240
Db 181 NEIFRHYDDFFERTILCNLSVMSCAVTRPGGLTYQEALESKAKQNLQSPPEPLIIPVL 240
Qy 241 YLTSILTHSRRLHEICDDIFAYVKORYFVEETVEIRNNGARLOCTILVLPSPHQNGFAN 300
Db 241 YLTSILTHSRRLHEICDDIFAYVKORYFVEETVEIRNNGARLOCTILVLPSPHQNGFAN 300
Qy 301 GHVNSVDGETIISDSDDSETQSCSFQNGKKDAIDPLLFYKVQOPTKKELHESAIVKAT 360
Db 301 GHVNSVDGETIISDSDDSETQSCSFQNGKKDAIDPLLFYKVQOPTKKELHESAIVKAT 360
Qy 361 QISRRKHLFSRDKLFLKQHCPEQGVIKIKASSLSYKIAEQDPSYFFPDPTTFIS 420
Db 361 QISRRKHLFSRDKLFLKQHCPEQGVIKIKASSLSYKIAEQDPSYFFPDPTTFIS 420
Qy 421 PANRRGRPPKRIHISQDNVANKOTLASYSKATKERDKLLKQEMKSLAFKAKLRE 480
Db 421 PANRRGRPPKRIHISQDNVANKOTLASYSKATKERDKLLKQEMKSLAFKAKLRE 480
Qy 481 KADALEAKKKEKDEKKREELKVIIEBERLKKBEKERLVEREKERLREBRKRYVE 540
Db 481 KADALEAKKKEKDEKKREELKVIIEBERLKKBEKERLVEREKERLREBRKRYVE 540
Qy 541 YLKQWSKPREDMECDDLKELPEPTPVKTRLPPEIFGDALMWLEFLNAGFELDLQDEFPD 600
Db 541 YLKQWSKPREDMECDDLKELPEPTPVKTRLPPEIFGDALMWLEFLNAGFELDLQDEFPD 600
Qy 601 GVTLEVLBEALVGNDSQGLCELLFFFLTAIFQAIABEEBAKEQLTDADTKGCSLSKL 660
Db 601 GVTLEVLBEALVGNDSQGLCELLFFFLTAIFQAIABEEBAKEQLTDADTKGCSLSKL 660
Qy 661 DLDSCTLSEILRLHLAGADVTSANAKRYQKRGFDATDADCMELSLNPSLVKKLSS 720
Db 661 DLDSCTLSEILRLHLAGADVTSANAKRYQKRGFDATDADCMELSLNPSLVKKLSS 720
Qy 721 TSVDYLTGPKGMKILHALCGKLLTLVSTRDFIEDYVDILRQAKQEFRELKAEQHRKERE 780
Db 721 TSVDYLTGPKGMKILHALCGKLLTLVSTRDFIEDYVDILRQAKQEFRELKAEQHRKERE 780
Qy 781 AAARIRKKEKELKEQEQMKKEQKLEKDEQRNSTADISIEBEREDFDTSIESKQTEQ 840
Db 781 AAARIRKKEKELKEQEQMKKEQKLEKDEQRNSTADISIEBEREDFDTSIESKQTEQ 840
Qy 841 KELDDMTFEDDDPGSHRGRGRGQNGKFEFTRQEQINCVTRRELLTADDEEALKQEH 900
Db 841 KELDDMTFEDDDPGSHRGRGRGQNGKFEFTRQEQINCVTRRELLTADDEEALKQEH 900
Qy 901 QRKELKELEKIQSAIACNTIIFPLGRDRMYRYWIFPSPGLFIEDYSGLTMEDMLLRPS 960
Db 901 QRKELKELEKIQSAIACNTIIFPLGRDRMYRYWIFPSPGLFIEDYSGLTMEDMLLRPS 960
Qy 961 SFQNVQSDQOVSTKTEGPELMSSESTSNIDQPRDHSVQLPKPVKPNRWCFYSSCEOLD 1020
Db 961 SFQNVQSDQOVSTKTEGPELMSSESTSNIDQPRDHSVQLPKPVKPNRWCFYSSCEOLD 1020
Qy 1021 QLIEALNBSRGHRESALKETLQEKSRICAOQLARFSEKHFSDKPPQPSKPTYSGRGSN 1080
Db 1021 QLIEALNBSRGHRESALKETLQEKSRICAOQLARFSEKHFSDKPPQPSKPTYSGRGSN 1080
Qy 1081 AYDPSQMAEQLELRLRDLFDLIEDRIYQGTGLGAIKVTDRHWRSALESGRYELLSEEN 1140
Db 1081 AYDPSQMAEQLELRLRDLFDLIEDRIYQGTGLGAIKVTDRHWRSALESGRYELLSEEN 1140

Qy 1141 KENGI IKTVNEDVEMEIDEQTKVIVKDRLLGIKTETPSTVSTNASTPQSSVSSVHYLAM 1200
Db 1141 KENGI IKTVNEDVEMEIDEQTKVIVKDRLLGIKTETPSTVSTNASTPQSSVSSVHYLAM 1200
Qy 1201 ALFOIEQGIERRFLKAPLDASDSGRSYKTVLDRWRESLLSSASLSQVFLHLSTLDRSVI 1260
Db 1201 ALFOIEQGIERRFLKAPLDASDSGRSYKTVLDRWRESLLSSASLSQVFLHLSTLDRSVI 1260
Qy 1261 SKSILNARCKICRKKGAENNVLCDDGDRGHHTYCVRPKLTVPBGDFWFCPCRPKQRCR 1320
Db 1261 SKSILNARCKICRKKGAENNVLCDDGDRGHHTYCVRPKLTVPBGDFWFCPCRPKQRCR 1320
Qy 1321 RLSFRQPSLESDESDMGDEVDGDEEGSQSEEEVEVEQDEDDSDSEEEVSLPK 1380
Db 1321 RLSFRQPSLESDESDMGDEVDGDEEGSQSEEEVEVEQDEDDSDSEEEVSLPK 1380
Qy 1381 RGRPQVRLPVKTRGKLSSSFSSRQOQEPGRYPERSQOQSTPKTTVSSKTGRSLAKINSAP 1440
Db 1381 RGRPQVRLPVKTRGKLSSSFSSRQOQEPGRYPERSQOQSTPKTTVSSKTGRSLAKINSAP 1440
Qy 1441 PTETKSLRIASRSTRHSHGPIQADVVELLSPPRRKRGRKSAANTPENSPPNPRVIAT 1500
Db 1441 PTETKSLRIASRSTRHSHGPIQADVVELLSPPRRKRGRKSAANTPENSPPNPRVIAT 1500
Qy 1501 KSSSQSRVNIASKLSIQESSEKRCRKROSPSPVTLGRSSRGQGVHLSAFEOLV 1560
Db 1501 KSSSQSRVNIASKLSIQESSEKRCRKROSPSPVTLGRSSRGQGVHLSAFEOLV 1560
Qy 1561 VELVRHDDSNPFLKLVSKIQVPDYDIIKKPIALNIIREKVNKCEYKLASEFIDDIELMF 1620
Db 1561 VELVRHDDSNPFLKLVSKIQVPDYDIIKKPIALNIIREKVNKCEYKLASEFIDDIELMF 1620
Qy 1621 SNCFEYNPRNTSEAKAGTRLOAFPHIOAQKLGHLHVTPSNVDQVSTPPAAKKSRI 1674
Db 1621 SNCFEYNPRNTSEAKAGTRLOAFPHIOAQKLGHLHVTPSNVDQVSTPPAAKKSRI 1674

RESULT 2

US-09-839-479-1
; Sequence 1, Application US/09839479
; Patent No. 672722
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 08501-042002
; CURRENT APPLICATION NUMBER: US/09/839, 479
; CURRENT FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1674
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-839-479-1

Query Match 100.0%; Score 8631; DB 4; Length 1674;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1674; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MEDASESRGVAPLINNVLPGLSPLSPVSVTGCKSHRVANKVEARSEKLLPTALPPSE 60
Db 1 MEDASESRGVAPLINNVLPGLSPLSPVSVTGCKSHRVANKVEARSEKLLPTALPPSE 60
Qy 61 PKVDOKLPRSSERRSGSGGTOPPARSRAVAAGEAARGAAGPERGSPGLGRVSPRCLCSG 120

Db 61 PKVDQKLPRSSRRSGGGTQPAKSRVAAGAAAGAPGRRVSPRLCSG 120
Qy 121 EGGQVAVGVIAGKRRGRGDSRRAPGGREMPLLHRKPFVRQKPADLRPDEVEFYCKVT 180
Db 121 EGGQVAVGVIAGKRRGRGDSRRAPGGREMPLLHRKPFVRQKPADLRPDEVEFYCKVT 180
Qy 181 NEIFRHYDDFPRTILCNLSVMSCAVTRPGGLTYQEALESEKAKQNTQSFPPELIIPVL 240
Db 181 NEIFRHYDDFPRTILCNLSVMSCAVTRPGGLTYQEALESEKAKQNTQSFPPELIIPVL 240
Qy 241 YLTSITHSRLHEICDDIFAYVKDYFVEETVEVIRNNCARLOCTILEVLPSSHONGAN 300
Db 241 YLTSITHSRLHEICDDIFAYVKDYFVEETVEVIRNNCARLOCTILEVLPSSHONGAN 300
Qy 301 GHVNSVDGETIISDSDSETQSCSFQNGKKDAIDPLLFKYKQVTKKELHESAIVRAT 360
Db 301 GHVNSVDGETIISDSDSETQSCSFQNGKKDAIDPLLFKYKQVTKKELHESAIVRAT 360
Qy 361 QISRRKHLFSRDKLFLKQHCPEQGVIKI KASSLSYTKIAEQDFSYFFDDPPTFFS 420
Db 361 QISRRKHLFSRDKLFLKQHCPEQGVIKI KASSLSYTKIAEQDFSYFFDDPPTFFS 420
Qy 421 PANRRGRPPKRIHISQEDNVANKOTLASYSRKATKRDKLLKQEMKSLAFKAKLRE 480
Db 421 PANRRGRPPKRIHISQEDNVANKOTLASYSRKATKRDKLLKQEMKSLAFKAKLRE 480
Qy 481 KADALEAKKKEDEKKEKREELKKTVEBERLKKKEKERLKVBERKEREKREERKKYVE 540
Db 481 KADALEAKKKEDEKKEKREELKKTVEBERLKKKEKERLKVBERKEREKREERKKYVE 540
Qy 541 YLKQWSKPREDMCEDDLKELPEPTPVKTRLPPEIFGDALMWLEFLNAGELFDLQDEFPD 600
Db 541 YLKQWSKPREDMCEDDLKELPEPTPVKTRLPPEIFGDALMWLEFLNAGELFDLQDEFPD 600
Qy 601 GVTLEVLREALVGNDSRGLPCELLFFFLTAIPQAIABEEVEVAKQOLTADOTKGCSSL 660
Db 601 GVTLEVLREALVGNDSRGLPCELLFFFLTAIPQAIABEEVEVAKQOLTADOTKGCSSL 660
Qy 661 DLDSCTLSEILRLHLASGADVTSANAKRYQKRGFGFATDACCMELRSLNSPLVKLSS 720
Db 661 DLDSCTLSEILRLHLASGADVTSANAKRYQKRGFGFATDACCMELRSLNSPLVKLSS 720
Qy 721 TSVYDLTPECKMKILHALCGKLLTLVSTRDFIEDYVDILRQAKQFRELKAEQHRKEREE 780
Db 721 TSVYDLTPECKMKILHALCGKLLTLVSTRDFIEDYVDILRQAKQFRELKAEQHRKEREE 780
Qy 781 AAARIRKKEEKLEKQEQKMKKEQKLEDEQORNSTADISIGEEBERDFDTSIESKQTEQ 840
Db 781 AAARIRKKEEKLEKQEQKMKKEQKLEDEQORNSTADISIGEEBERDFDTSIESKQTEQ 840
Qy 841 KELDDQMTFEDDDPGSHKRRGRGKRGONGKFEFTROQINCVTRRELTADDEEALKQEH 900
Db 841 KELDDQMTFEDDDPGSHKRRGRGKRGONGKFEFTROQINCVTRRELTADDEEALKQEH 900
Qy 901 QRKEKELLEKIQSAIACNIPGLGRDRMYRRIWIPPSIPGLFIEDYSGLTEDMLLRPS 960
Db 901 QRKEKELLEKIQSAIACNIPGLGRDRMYRRIWIPPSIPGLFIEDYSGLTEDMLLRPS 960
Qy 961 SFQNNVQSQDPQVSTKTGEPLMSESTSNIDQPRDHSVQLPKPVHKPNRWCFYSSCEOLD 1020
Db 961 SFQNNVQSQDPQVSTKTGEPLMSESTSNIDQPRDHSVQLPKPVHKPNRWCFYSSCEOLD 1020
Qy 1021 QLIEALNRSRGHRESALKETLLQEKSRICAQLARFSEKPFHSDKQPKDPSKPTYSRGRSN 1080
Db 1021 QLIEALNRSRGHRESALKETLLQEKSRICAQLARFSEKPFHSDKQPKDPSKPTYSRGRSN 1080
Qy 1081 AYDPSQMAEKOLELRDLFLDDIEDRIYQGTGLGAKVTDRHIWRSALSGRYELLSEN 1140
Db 1081 AYDPSQMAEKOLELRDLFLDDIEDRIYQGTGLGAKVTDRHIWRSALSGRYELLSEN 1140
Qy 1141 KENGIITKTVNEDVEEMEIDEQTKVIVKORLLGIKTETPSTVSTNASTPQSVSSVVHYLAM 1200

Db 1141 KENGIITKTVNEDVEEMEIDEQTKVIVKORLLGIKTETPSTVSTNASTPQSVSSVVHYLAM 1200
Qy 1201 ALFQIEQGIERRFLKAPLDASDSGRSYKTVLDRWRRESLLSSASLSQVFLHLSTLDRSVTW 1260
Db 1201 ALFQIEQGIERRFLKAPLDASDSGRSYKTVLDRWRRESLLSSASLSQVFLHLSTLDRSVTW 1260
Qy 1261 SKSILNARCKICRKKGAENMWLDCGCDRGHHITYCVRPKLKTVPBGDWFCPCRCRQRCR 1320
Db 1261 SKSILNARCKICRKKGAENMWLDCGCDRGHHITYCVRPKLKTVPBGDWFCPCRCRQRCR 1320
Qy 1321 RLSFRQPSLESDEVEDSMGDEDEVGDDEEGOSBEEVEVEDEDDSQEVEEVSPLK 1380
Db 1321 RLSFRQPSLESDEVEDSMGDEDEVGDDEEGOSBEEVEVEDEDDSQEVEEVSPLK 1380
Qy 1381 RGRPOVRLPVKTRGKLSSSFSSRGOQEPGRYPSRSQOSTPKTTVSSKTGRSLRKINSAP 1440
Db 1381 RGRPOVRLPVKTRGKLSSSFSSRGOQEPGRYPSRSQOSTPKTTVSSKTGRSLRKINSAP 1440
Qy 1441 PTETKSLRIASRSTRSHGPIQADVVELLSPPRRKRGRKKSANNTPENSPPNFRVIAT 1500
Db 1441 PTETKSLRIASRSTRSHGPIQADVVELLSPPRRKRGRKKSANNTPENSPPNFRVIAT 1500
Qy 1501 KSSQOSRVSNTASKLSLOESKRCRCRQSPSPVTLGRSSGRCQGVHLSAPEQLV 1560
Db 1501 KSSQOSRVSNTASKLSLOESKRCRCRQSPSPVTLGRSSGRCQGVHLSAPEQLV 1560
Qy 1561 VELVRHDSWPFLKLVSKIQVDDYDIKKPIALNIIREKVNKCEYKLASEFIDDIELMF 1620
Db 1561 VELVRHDSWPFLKLVSKIQVDDYDIKKPIALNIIREKVNKCEYKLASEFIDDIELMF 1620
Qy 1621 SNCFEYNPRNTSEAKAGTRLOAFHHIOAKLGLHVTSPNVDQVSTPPAAKKSRI 1674
Db 1621 SNCFEYNPRNTSEAKAGTRLOAFHHIOAKLGLHVTSPNVDQVSTPPAAKKSRI 1674

RESULT 3
US-09-418-710-70
; Sequence 70, Application US/09418710
; Patent No. 6596482
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042001
; CURRENT APPLICATION NUMBER: US/09/418,710
; CURRENT FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 70
; LENGTH: 1673
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-418-710-70

Query Match 99.5%; Score 8588.5; DB 4; Length 1673;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 1667; Conservative 3; Mismatches 3; Indels 1; Gaps 1;

Qy 1 MEDASESRGVAPLINNVLPGSLPLSVTVGCKSHRVANKVEARSEKLLPTALPSE 60
Db 1 MEDASESRGVAPLINNVLPGSLPLSVTVGCKSHRVANKVEARSEKLLPTALPSE 60
Qy 61 PKVDQKLPRSSRRSGGGTQPAKSRVAAGAAAGAPGRRVSPRLCSG 120
Db 61 PKVDQKLPRSSRRSGGGTQPAKSRVAAGAAAGAPGRRVSPRLCSG 120
Qy 121 EGGQVAVGVIAGKRRGRGDSRRAPGGREMPLLHRKPFVRQKPADLRPDEVEFYCKVT 180
Db 121 EGGQVAVGVIAGKRRGRGDSRRAPGGREMPLLHRKPFVRQKPADLRPDEVEFYCKVT 180
Qy 181 NEIFRHYDDFPRTILCNLSVMSCAVTRPGGLTYQEALESEKAKQNTQSFPPELIIPVL 240
Db 181 NEIFRHYDDFPRTILCNLSVMSCAVTRPGGLTYQEALESEKAKQNTQSFPPELIIPVL 240
Qy 241 YLTSITHSRLHEICDDIFAYVKDYFVEETVEVIRNNCARLOCTILEVLPSSHONGAN 300
Db 241 YLTSITHSRLHEICDDIFAYVKDYFVEETVEVIRNNCARLOCTILEVLPSSHONGAN 300
Qy 301 GHVNSVDGETIISDSDSETQSCSFQNGKKDAIDPLLFKYKQVTKKELHESAIVRAT 360
Db 301 GHVNSVDGETIISDSDSETQSCSFQNGKKDAIDPLLFKYKQVTKKELHESAIVRAT 360
Qy 361 QISRRKHLFSRDKLFLKQHCPEQGVIKI KASSLSYTKIAEQDFSYFFDDPPTFFS 420
Db 361 QISRRKHLFSRDKLFLKQHCPEQGVIKI KASSLSYTKIAEQDFSYFFDDPPTFFS 420
Qy 421 PANRRGRPPKRIHISQEDNVANKOTLASYSRKATKRDKLLKQEMKSLAFKAKLRE 480
Db 421 PANRRGRPPKRIHISQEDNVANKOTLASYSRKATKRDKLLKQEMKSLAFKAKLRE 480
Qy 481 KADALEAKKKEDEKKEKREELKKTVEBERLKKKEKERLKVBERKEREKREERKKYVE 540
Db 481 KADALEAKKKEDEKKEKREELKKTVEBERLKKKEKERLKVBERKEREKREERKKYVE 540
Qy 541 YLKQWSKPREDMCEDDLKELPEPTPVKTRLPPEIFGDALMWLEFLNAGELFDLQDEFPD 600
Db 541 YLKQWSKPREDMCEDDLKELPEPTPVKTRLPPEIFGDALMWLEFLNAGELFDLQDEFPD 600
Qy 601 GVTLEVLREALVGNDSRGLPCELLFFFLTAIPQAIABEEVEVAKQOLTADOTKGCSSL 660
Db 601 GVTLEVLREALVGNDSRGLPCELLFFFLTAIPQAIABEEVEVAKQOLTADOTKGCSSL 660
Qy 661 DLDSCTLSEILRLHLASGADVTSANAKRYQKRGFGFATDACCMELRSLNSPLVKLSS 720
Db 661 DLDSCTLSEILRLHLASGADVTSANAKRYQKRGFGFATDACCMELRSLNSPLVKLSS 720
Qy 721 TSVYDLTPECKMKILHALCGKLLTLVSTRDFIEDYVDILRQAKQFRELKAEQHRKEREE 780
Db 721 TSVYDLTPECKMKILHALCGKLLTLVSTRDFIEDYVDILRQAKQFRELKAEQHRKEREE 780
Qy 781 AAARIRKKEEKLEKQEQKMKKEQKLEDEQORNSTADISIGEEBERDFDTSIESKQTEQ 840
Db 781 AAARIRKKEEKLEKQEQKMKKEQKLEDEQORNSTADISIGEEBERDFDTSIESKQTEQ 840
Qy 841 KELDDQMTFEDDDPGSHKRRGRGKRGONGKFEFTROQINCVTRRELTADDEEALKQEH 900
Db 841 KELDDQMTFEDDDPGSHKRRGRGKRGONGKFEFTROQINCVTRRELTADDEEALKQEH 900
Qy 901 QRKEKELLEKIQSAIACNIPGLGRDRMYRRIWIPPSIPGLFIEDYSGLTEDMLLRPS 960
Db 901 QRKEKELLEKIQSAIACNIPGLGRDRMYRRIWIPPSIPGLFIEDYSGLTEDMLLRPS 960
Qy 961 SFQNNVQSQDPQVSTKTGEPLMSESTSNIDQPRDHSVQLPKPVHKPNRWCFYSSCEOLD 1020
Db 961 SFQNNVQSQDPQVSTKTGEPLMSESTSNIDQPRDHSVQLPKPVHKPNRWCFYSSCEOLD 1020
Qy 1021 QLIEALNRSRGHRESALKETLLQEKSRICAQLARFSEKPFHSDKQPKDPSKPTYSRGRSN 1080
Db 1021 QLIEALNRSRGHRESALKETLLQEKSRICAQLARFSEKPFHSDKQPKDPSKPTYSRGRSN 1080
Qy 1081 AYDPSQMAEKOLELRDLFLDDIEDRIYQGTGLGAKVTDRHIWRSALSGRYELLSEN 1140
Db 1081 AYDPSQMAEKOLELRDLFLDDIEDRIYQGTGLGAKVTDRHIWRSALSGRYELLSEN 1140
Qy 1141 KENGIITKTVNEDVEEMEIDEQTKVIVKORLLGIKTETPSTVSTNASTPQSVSSVVHYLAM 1200

Db 121 EGGQVAVGVIAKGRGRRGDSRRAPGGRGEMPLLLHRKPFVRQKPPADLRPDEEVFYCKVT 180
Qy 181 NEIFRHYDDFERITILCNLSVWSCAVTCRPGLTYYQEALESEKKARONLQSPPEPLIIPVL 240
Db 181 NEIFRHYDDFERITILCNLSVWSCAVTCRPGLTYYQEALESEKKARONLQSPPEPLIIPVL 240
Qy 241 YLTSILTHSRSLHEICDDIFAYVKDORYFVEETVEVIRNNGARLQCTILEVLPPSHONGFAN 300
Db 241 YLTSILTHSRSLHEICDDIFAYVKDORYFVEETVEVIRNNGARLQCTILEVLPPSHONGFAN 300
Qy 301 GHNVSDGETIISDSDDSETQSCSFQNGKKDAIDPLLFYKYQVPTKKELHESAIVKAT 360
Db 301 GHNVSDGETIISDSDDSETQSCSFQNGKKDAIDPLLFYKYQVPTKKELHESAIVKAT 360
Qy 361 QISRKHLPFRDCLKLFLKQCEPQEGVIKIKASSLSYTKIABODFSYFFPDPTTIFS 420
Db 361 QISRKHLPFRDCLKLFLKQCEPQEGVIKIKASSLSYTKIABODFSYFFPDPTTIFS 420
Qy 421 PANRRRGRPPKRIHISQBDNVANKOTLASYSKATKRDKLLKOEEMKSLAFKAKLKRE 480
Db 421 PANRRRGRPPKRIHISQBDNVANKOTLASYSKATKRDKLLKOEEMKSLAFKAKLKRE 480
Qy 481 KADALEAKKSKEDKKEKREBKXIVEBERLKKKEERLKVBEREKREKLREKRYVE 540
Db 481 KADALEAKKSKEDKKEKREBKXIVEBERLKKKEERLKVBEREKREKLREKRYVE 540
Qy 541 YLKOWSKREDMECDLDELPEPTPVKTRLPPELFGDALMVLEFLNAPGELFDLODEPPD 600
Db 541 Y- KOWSKREDMECDLDELPEPTPVKTRLPPELFGDALMVLEFLNAPGELFDLODEPPD 599
Qy 601 GVTLEVLAEALVGNDSQPLCELLFFLTAIFQALAEBEVEVAKEQLTDADTKGCSLKL 660
Db 600 GVTLEVLAEALVGNDSQPLCELLFFLTAIFQALAEBEVEVAKEQLTDADTKGCSLKL 659
Qy 661 DLDSCTLSEILRLHILASGADVTSANAKYRQKRGGFATDDACMELRSLNPSLVKKLSS 720
Db 660 DLDSCTLSEILRLHILASGADVTSANAKYRQKRGGFATDDACMELRSLNPSLVKKLSS 719
Qy 721 TSVVDLTPGEQKILHALCGKLLTLVSTRDIEDYVDILROAKQEFRELKAEQHRKREE 780
Db 720 TSVVDLTPGEQKILHALCGKLLTLVSTRDIEDYVDILROAKQEFRELKAEQHRKREE 779
Qy 781 AAARIRKKEKLEKEQEKMEKEKLEKDEQORNSTADISIGEEEREDFDTISIESKQTEQ 840
Db 780 AAARIRKKEKLEKEQEKMEKEKLEKDEQORNSTADISIGEEEREDFDTISIESKQTEQ 839
Qy 841 KELDQDMFTEDDDPGSHKRGRRGKRGQNGKFETRQEQINCVTRELLTADEEALQKEH 900
Db 840 KELDQDMFTEDDDPGSHKRGRRGKRGQNGKFETRQEQINCVTRELLTADEEALQKEH 899
Qy 901 QRKEKELLEKQSAIACTNIIFPLGRDRMYRYWIFFPSIPGLFIEDYSGLTEDMLLRPS 960
Db 900 QRKEKELLEKQSAIACTNIIFPLGRDRMYRYWIFFPSIPGLFIEDYSGLTEDMLLRPS 959
Qy 961 SFQNVNVSQDPQVSTKTGEPLMSSTSNIDGPRDHSVQLPKPVHKPNRWCFYSSCQOLD 1020
Db 960 SFQNVNVSQDPQVSTKTGEPLMSSTSNIDGPRDHSVQLPKPVHKPNRWCFYSSCQOLD 1019
Qy 1021 QLIIEALNSRGHRESALKETLLQESKRICAQLARFSEKFFHPSDKPQDPKPTYSRGRSSN 1080
Db 1020 QLIIEALNSRGHRESALKETLLQESKRICAQLARFSEKFFHPSDKPQDPKPTYSRGRSSN 1079
Qy 1081 AYDPSQMAEQLELRDLFLDIEDRIYQGTGLAIVKVTDRHIWRSALSGRYELLSEEN 1140
Db 1080 AYDPSQMAEQLELRDLFLDIEDRIYQGTGLAIVKVTDRHIWRSALSGRYELLSEEN 1139
Qy 1141 KENGIITKTVNEDVEMEIDEOTKIVKDRLLAGITETPSTVSTNASTPQSVSSVVHYLAM 1200
Db 1140 KENGIITKTVNEDVEMEIDEOTKIVKDRLLAGITETPSTVSTNASTPQSVSSVVHYLAM 1199
Qy 1201 ALFOIEQGIERRFLKAPLDASDSGRSYKTVLDRWRESLSSASLSQVFLHSLTLDRSVIW 1260
Db 1200 ALFOIEQGLERRFLKAPLDASDSGRSYKTVLDRWRESLSSASLSQVFLHSLTLDRSVIW 1259

Qy 1261 SKSILNARCKICRKKGDAENNVLCDCGDRGHHTYCVRPKLKTVPEGDWFCPECRPKQRCR 1320
Db 1260 SKSILNARCKICRKKGDAENNVLCDCGDRGHHTYCVRPKLKTVPEGDWFCPECRPKQRCR 1319
Qy 1321 RLSFRQPSLESDESDVEDSMGDEDEVDGDEEQSEBEEYEVEDEQDDSOEEBESVLPK 1380
Db 1320 RLSFRQPSLESDESDVEDSMGDEDEVDGDEEQSEBEEYEVEDEQDDSOEEBESVLPK 1379
Qy 1381 RGRPOVRLPVKTRKGLSSSFSSRQGOQRPGRYPSSRSQOSTPKTTVSSKTGRSLRKINSAP 1440
Db 1380 RGRPOVRLPVKTRKGLSSSFSSRQGOQRPGRYPSSRSQOSTPKTTVSSKTGRSLRKINSAP 1439
Qy 1441 PTEKTSRLIASRSTRHSHGPLQADVVELLSPREKRRGRKSANNTPENSFPNFRVIAT 1500
Db 1440 PTEKTSRLIASRSTRHSHGPLQADVVELLSPREKRRGRKSANNTPENSFPNFRVIAT 1499
Qy 1501 KSSEQSRSVNIASKLSLQESKRCRKRQSPSPVTLGRSSRGQGVHLSAFEOLV 1560
Db 1500 KSSEQSRSVNIASKLSLQESKRCRKRQSPSPVTLGRSSRGQGVHLSAFEOLV 1559
Qy 1561 VELVRHDDSWPFLKLVSKIQVPDYDIIKKPIALNIIREKVNKCEYKLASEFIDDIELMF 1620
Db 1560 VELVRHDDSWPFLKLVSKIQVPDYDIIKKPIALNIIREKVNKCEYKLASEFIDDIELMF 1619
Qy 1621 SNCFEYNPRNTSEAKAGTRLOAFPHIOAKLGLHVTPSNVDOVSTPPAAKKSRI 1674
Db 1620 SNCFEYNPRNTSEAKAGTRLOAFPHIOAKLGLHVTPSNVDOVSTPPAAKKSRI 1673

RESULT 4

US-09-839-479-69
; Sequence 69, Application US/09839479
; Patent No. 6727222
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042002
; CURRENT APPLICATION NUMBER: US/09/839,479
; CURRENT FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 69
; LENGTH: 1673
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-839-479-69

Query Match 99.5%; Score 8588.5; DB 4; Length 1673;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 1667; Conservative 3; Mismatches 3; Indels 1; Gaps 1;

Qy 1 MEDASESSRGVAPLINNVNVLPGSPLSLPVSVTGCKSHRVANKVYAEARSEKLLPTALPPSE 60
Db 1 MEDASESSRGVAPLINNVNVLPGSPLSLPVSVTGCKSHRVANKVYAEARSEKLLPTALPPSE 60
Qy 61 PKVDQKLPRSSRRGSGGQTQFPARSRAVAAGAAAGAPRGSPILGRVSPRCLCSG 120
Db 61 PKVDQKLPRSSRRGSGGQTQFPARSRAVAAGAAAGAPRGSPILGRVSPRCLCSG 120
Qy 121 EGGQVAVGVIAKGRGRRGDSRRAPGGRGEMPLLLHRKPFVRQKPPADLRPDEEVFYCKVT 180
Db 121 EGGQVAVGVIAKGRGRRGDSRRAPGGRGEMPLLLHRKPFVRQKPPADLRPDEEVFYCKVT 180
Qy 181 NEIFRHYDDFERITILCNLSVWSCAVTCRPGLTYYQEALESEKKARONLQSPPEPLIIPVL 240

181	Db	 NEIFHYDDFERTILCNLSVMSCAVTGRGLTYQEALESSEKQARONLQSPPEPLIIPVL	241
241	Qy	YLTSLTHRSRLHEICDDIFAYVKORYFYVEETVEVIRNNGARLOCTILVLPPSHONGFAN	300
241	Db	YLTSLTHRSRLHEICDDIFAYVKORYFYVEETVEVIRNNGARLOCTILVLPPSHONGFAN	300
301	Qy	GHVNSVDGETIIISDSDDSETQSCSFQNGKKKDAIDPLLFYKYQOPTKELHESAIVKAT	360
301	Db	GHVNSVDGETIIISDSDDSETQSCSFQNGKKKDAIDPLLFYKYQOPTKELHESAIVKAT	360
361	Qy	QISRRKHLFSDKULKLFUKOCEPQEGVIKIKASSLSYTKIABQDFSVFFPDDPPTIFS	420
361	Db	QISRRKHLFSDKULKLFUKOCEPQEGVIKIKASSLSYTKIABQDFSVFFPDDPPTIFS	420
421	Qy	PANRRRRGPPRRKIHISQEDNVANKOTLASYESKATKRDKLLKOBEMKSLAFKAKUKRE	480
421	Db	PANRRRRGPPRRKIHISQEDNVANKOTLASYESKATKRDKLLKOBEMKSLAFKAKUKRE	480
481	Qy	KADALEAKKKEKEDKKEKREELKKIVEBERLKKKEERLKVBEREKREKLREERKRYVE	540
481	Db	KADALEAKKKEKEDKKEKREELKKIVEBERLKKKEERLKVBEREKREKLREERKRYVE	540
541	Qy	YLKQWSXPREDMECDLDELPEPTPVKTRLDPETIFGDALMVLEFLNAPGELFDLODFPD	600
541	Db	Y - KQWSXPREDMECDLDELPEPTPVKTRLDPETIFGDALMVLEFLNAPGELFDLODFPD	599
601	Qy	GVTLEVLAEALVGNDSGPELCELLFFFTLTAIQAIAEBEEBEVAKQEQLTADTKGCSLKL	660
600	Db	GVTLEVLAEALVGNDSGPELCELLFFFTLTAIQAIAEBEEBEVAKQEQLTADTKGCSLKL	659
661	Qy	DLDSCTLSEILRLHLIAGSDVTSANAKRYQKRGGFDDADACMELRSLNPSLVKKLSS	720
660	Db	DLDSCTLSEILRLHLIAGSDVTSANAKRYQKRGGFDDADACMELRSLNPSLVKKLSS	719
721	Qy	TSVVDLTPGCKMILHALCGKLLTLVSTRDPIEDYVDILROAKQEFRELKABQHRKREE	780
720	Db	TSVVDLTPGCKMILHALCGKLLTLVSTRDPIEDYVDILROAKQEFRELKABQHRKREE	779
781	Qy	AAAIRKRKBEKLEQEOQKMEKQELKEDBORNSTADISIGEEREDFDTSIESKQTEQ	840
780	Db	AAAIRKRKBEKLEQEOQKMEKQELKEDBORNSTADISIGEEREDFDTSIESKQTEQ	839
841	Qy	KELQDMFTEDDDPGSHKRRGRKRGONGKFETROEQINCVTRELLTADDEEALQEH	900
840	Db	KELQDMFTEDDDPGSHKRRGRKRGONGKFETROEQINCVTRELLTADDEEALQEH	899
901	Qy	QRKEKELLEKIQSAIACTNIPPLGRDRMYRYWIFFPSIPGLFIBEDYSGLTEDMLLRPPS	960
900	Db	QRKEKELLEKIQSAIACTNIPPLGRDRMYRYWIFFPSIPGLFIBEDYSGLTEDMLLRPPS	959
961	Qy	SFQNNVQSDPQVSTKTGPELWMSSTSNIDGPRDHSVOLPKPVHKPNRWCIFYSSCOLD	1020
960	Db	SFQNNVQSDPQVSTKTGPELWMSSTSNIDGPRDHSVOLPKPVHKPNRWCIFYSSCOLD	1019
1021	Qy	QLIEALNSRGHRESALKETLLQEKSRICAQLARFSEBKHFSDPKQDPSKPTYSRGRSSN	1080
1020	Db	QLIEALNSRGHRESALKETLLQEKSRICAQLARFSEBKHFSDPKQDPSKPTYSRGRSSN	1079
1081	Qy	AYDPSQWCAEQBLERLDFLLDIEDRIYQOTLGAIKVTDRIHWRSALESGRYELLSEEN	1140
1080	Db	AYDPSQWCAEQBLERLDFLLDIEDRIYQOTLGAIKVTDRIHWRSALESGRYELLSEEN	1139
1141	Qy	KENGIIKTIVNEDVEMEIDEQTKVIVKDRLLGIIKTETPTSTVSTWASTPQSVSSVVHYLAM	1200
1140	Db	KENGIIKTIVNEDVEMEIDEQTKVIVKDRLLGIIKTETPTSTVSTWASTPQSVSSVVHYLAM	1199
1201	Qy	ALFOIEOGIERRRFLKAPLDASDSGRSYKTVLDRWRSELLSSASLSQVFLHLSTLDRSVI	1260
1200	Db	ALFOIEOGLERRFLKAPLDASDSGRSYKTVLDRWRSELLSSASLSQVFLHLSTLDRSVI	1259
1261	Qy	SKSILNARCKICRKKGDAENMVLCGDGRGHHTYCVRPKLTVPBGDMWFCPECRPKQCR	1320

Db	1260	SKSILNARCKIKRCKKGDAENWVLCDGCDRGHHTYVVRPKLKIIVPGDWFCEPCRCRQRCR	1319
Qy	1321	RLSFRQPSLESDEBVEDSDMSGGDEVDGDEEGQGESEEEYEVQEDEDSDQEEEEVSLPK	1380
Db	1320	RLSFRQPSLESDEBVEDSDMSGGDEVDGDEEGQGESEEEYEVQEDEDSDQEEEEVSLPK	1379
Qy	1381	RGRPQVRLPVKTRGKLSFSSSRGQQQEPGRYPSRSQQSTPKTTVSSKTGRSLURKINSAP	1440
Db	1380	RGRPQVRLPVKTRGKLSFSSSRGQQQEPGRYPSRSQQSTPKTTVSSKTGRSLURKINSAP	1439
Qy	1441	PTTETKSURIASRSTRSHGHPLOADVVFELLSPRRKRGKRSANNTPENSNNFNFVFIAT	1500
Db	1440	PTTETKSURIASRSTRSHGHPLOADVVFELLSPRRKRGKRSANNTPENSNNFNFVFIAT	1499
Qy	1501	KSEQGRSVNIAKSLSLQESKSRCKRKQSPSPVTLGRRSSGGQGVHLELSAFEQLV	1560
Db	1500	KSEQGRSVNIAKSLQESKSRCKRKQSPSPVTLGRRSSGGQGVHLELSAFEQLV	1559
Qy	1561	VELVRHDDSPFPLKLVSKIQVPDYDIIKKPIALNIIREKVNKCEYKLASEFIDDIELMF	1620
Db	1560	VELVRHDDSPFPLKLVSKIQVPDYDIIKKPIALNIIREKVNKCEYKLASEFIDDIELMF	1619
Qy	1621	SNCFEYNPRNTSBAKAGTRLQAFPHIQAOKGLGHVTPNSNDQVSTVPAAKKSRI	1674
Db	1620	SNCFEYNPRNTSBAKAGTRLQAFPHIQAOKGLGHVTPNSNDQVSTVPAAKKSRI	1673

```

RESULT 5
US-09-418-710-63
; Sequence 63, Application US/09418710
; Patent No. 6596482
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042001
; CURRENT APPLICATION NUMBER: US/09/418, 710
; CURRENT FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 63
; LENGTH: 211
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-418-710-63

```

Query Match	12.6%	Score 1087;	DB 4;	Length 211;
Best Local Similarity	100.0%;	Pred. No. 3.2e-72;		
Matches 211; Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;	
Qy	532	REEKRYVEYLKOWSKPREDMECDLXELPEPVPVKTRLPPEIFGDALMWLEFLNAFCEL	591	
Db	1	REEKRYVEYLKOWSKPREDMECDLXELPEPVPVKTRLPPEIFGDALMWLEFLNAFCEL	60	
Qy	592	FDLQDFPDGVTLEVEEALVGNDSGEPGLCELLAFFLTAIQAIAESEEVAKEQLTDAD	651	
Db	61	FDLQDFPDGVTLEVEEALVGNDSGEPGLCELLAFFLTAIQAIAESEEVAKEQLTDAD	120	
Qy	652	TKGCSLKSLLDSCSTLSEIILRLHLHLAGSADVTSANAKRYQKRGCFPATDDACMELRLSN	711	
Db	121	TKGCSLKSLLDSCSTLSEIILRLHLHLAGSADVTSANAKRYQKRGCFPATDDACMELRLSN	180	
Qy	712	PSLVKXLSSTSVYDLTPGERKWKILHALCGKL	742	
Db	181	PSLVKXLSSTSVYDLTPGERKWKILHALCGKL	211	

RESULT 6

```
US-09-839-479-62
; Sequence 62, Application US/09839479
; Patent No. 6727222
; GENERAL INFORMATION:
; APPLICANT: JONES, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042002
; CURRENT APPLICATION NUMBER: US/09/839,479
; CURRENT FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 62
; LENGTH: 211
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-839-479-62

Query Match 12.6%; Score 1087; DB 4; Length 211;
Best Local Similarity 100.0%; Pred. No. 3.2e-72;
Matches 211; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 532 REEKRYVEYLKQSKPREDMECDLKPPEPTPVKTRLPPEIFGDALMWLEFLNAGEL 591
Db 1 REEKRYVEYLKQSKPREDMECDLKPPEPTPVKTRLPPEIFGDALMWLEFLNAGEL 60

Qy 592 FDLQDEFFPDGVTLEVEALVGNDSSEGLPCELLFFFLTAIFQAIABEEVEVAKSOLTDAD 651
Db 61 FDLQDEFFPDGVTLEVEALVGNDSSEGLPCELLFFFLTAIFQAIABEEVEVAKSOLTDAD 120

Qy 652 TKGCSLSLSDSCTLSILRHILHILASGADVTSANAKRYQKRGGFDTDDACMELRLSN 711
Db 121 TKGCSLSLSDSCTLSILRHILHILASGADVTSANAKRYQKRGGFDTDDACMELRLSN 180

Qy 712 PSLVKLSSTSVYDLTPGCKMKILHALCGKL 742
Db 181 PSLVKLSSTSVYDLTPGCKMKILHALCGKL 211

RESULT 7
US-09-949-016-7037
; Sequence 7037, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7037
; LENGTH: 1540
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-7037

Query Match 11.1%; Score 954.5; DB 4; Length 1540;
Best Local Similarity 23.5%; Pred. No. 3.7e-61;
```

```
Matches 419; Conservative 264; Mismatches 556; Indels 547; Gaps 70;
Qy 86 SRAVAAGEAAARGAA---GPERGSPLGRRVSPRLCSGEGQVAVGIAGKRRGRDGS 142
Db 2 SRGAGGEGAGVRAALSPLSPPTPPGAWRSR-----AAGPSAAPRGPRQPPV 51
Qy 143 RRAPGGREMPLLHRRKPFVROKPPADLRPDDEBFYCKVTNEIFRHYDFFETILCNSLVW 202
Db 52 PTAAAM-APLLGRKPPFLVKP---LPGEPLFTIPHITQEAFTREEVEARLEYSERIW 107
Qy 203 SCAVTRPGLTYQEALESEKKAQNL-QSPF---EPLIPLVLYLTSTHRSRLHEICDDI 258
Db 108 TCKSTGSSQLTHKEAWEEQEVALLKEEPFAWYELVLEMVH---HNTASLEKLVDTA 163
Qy 259 PAYVKDRYFVEETVEVRNNGCARLOCTILEVLP-----PS-----HQ 295
Db 164 WLEIMTRYAVGEECDFEVGEKMKLVKIVKIHPLEKVEDEATEKSPDGDSPSSDKENS 223
Qy 296 NGFANGHVNSVDGETIISDSDDSETOS-----CSFONGKKKDAIDPLL-FKYK 343
Db 224 SOIAQDHQK---ETVVKEDGREGRESINDRARRSPRKLPTSLKKGKRWAPKFLPHKYD 280
Qy 344 VQPTKELHESAI---VKATQISRRKHLFSRDKLFLKQHC-----EPOEGVIK---I 391
Db 281 V---KLQNEKDIISNVPADSLIRTERPPNKEIVRYEIRHNALRAGTGENAPWVVEDELV 336
Qy 392 KASLSITYKTAEQDFSFYFFDDPPTFFPSPANRRG-----RPPKRIHISQEDN----- 440
Db 337 KKYSLPS-----KFSDFLDPYKYMTLNFSTKTKNGTSPDRKPSK---SKTDNSSLSS 387
Qy 441 -----VANKOTLA-----SVRSKATKER----- 458
Db 388 PLNPKLVCHVHLKXKLSGSPKLVKXNSKNSKSPPEHLEEMKOMSPNKLTNFTHPKGGPP 447
Qy 459 -----DKLLK-----QEEKSLAFKAKLKR----- 479
Db 448 AKPGKHSKPLKAKGRSKGILNGQKSTGNSKSPKGLKTPKTKWKQWTLMDMAKGTKQM 507
Qy 480 -----EKADALEAKKKEKEDYEKKR----- 499
Db 508 TRAPRNSGGTPTSSKPHKHLPPAALHLIAYYKENKDRKREDKRSALSCVISKARLLSSED 567
Qy 500 -----EELKKIVBEE-----RLKKKEERKLVKEREKEREKLR--EKKRYVEY 541
Db 568 RARLPEELSLVQKRYELLEHKKXWASMSBEQRKEYLKKXREELKKLKEKAKERREKEM 627
Qy 542 LKOWSKPR--EDMECDLKPPEPTPVKT--RLPPEIFGDALMWLEFLNAGELFDLQDE 597
Db 628 LERLEKQRYEDQELTG-KNLPAFLVDTPBGLPNTLFGDVAMVVEFLSCYSGLLLPDAQ 686
Qy 598 FPDGVTLEVEALVGNDSSEGLPCELLFFFLTAIFQAIABEE--EEVAKSOLTDADTKGC 655
Db 687 YP--ITAVSLMEAL-SADKGG-----FLYLNRLVILLOTLLQDEIAE----DYGELGM 733
Qy 656 SLKSLDSDSCTLSILRHILHILASGADVTSANAKRYQKRGGFDTDDACMELRLSNFSLV 715
Db 734 KLSEIPLTHSVSELVRLCLRRSDVQEESESGD-----DNKDSAAFEDNEVQDEFL 786
Qy 716 KKLSTSVYDLTPGCKMKILHALCGKLLTLVSTRDFIEDYVDILRQAKQEFRELKAEQHR 775
Db 787 EKLTSBFFELTSEKLIQILTALCHRILMTYVQD-----HM 823
Qy 776 KEREAAAARIRKKEEKLKEQEQKKEKQEKLEDE---QBNSTADISIGEEEREDPDS 832
Db 824 ETRQMGAEIWKERLAVLKEENDKKRAEKKRKEAMEANKENGKVGGLGKTRK----- 878
Qy 833 IESKDTSEKSLDQDMFTEDDDPGSHKGR-----RGKRGQNGKPFETRQEQINCVTRELL 888
Db 879 -----KEIVKFPQVDTEADMISAVKSRLLAIQAK-----KERIQE-----REMK 921
Qy 889 TADEEALKQEHORKEKELLEK-IQSAIACTNIF-----PLGRDRMYRYWIF-PSIPGLF 942
Db 922 VKLERQA-EEERIRKHAARAKAFQEGIAKAKLVMRRTPIGTDRNHNRYMLFSDVPLGF 980
```



```
QY 1146 IKTV--NBDVEMEIDQTKVIVKDRLLGIKTKTPTSTVSTWASTPQSVS----- 1192
: : ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1583 TRDIVENWSVEEQAMENDLSVLQOVEDLERRVVASASLQVKGWMCPEPASEREDLVYFEHK 1642
: : ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1193 -----SVVHYL-----AMALFOIEOGIERRFUK--APLDAS 1221
: : ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1643 SFTKCKEHGDEFTGEDESSAHLERKSDNPLDIAVTRLADLERNIERIEDIAP---- 1698
: : ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1222 DSGRSYKTVLDRWRRESLLSSASLQVFLHLSTLDRSVIWSKSIINARCKICRKGDAENM 1281
: : ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1699 -----GLVRWRLSEARSAQVALCIIQLOLQKSIWAKESIMKVYQICRKGDNELL 1750
: : ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1282 VLDGCDGRHRTYCVRPKLTVPBGDMFPCPCRPKQ-----RCRLSPRQPSLESDEDV 1336
: : ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1751 LLDGCDKGCHTYCHRPKITIPDGMFCPACIAKASQTLKIKLHVKGKKTWESKK-- 1808
: : ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1337 EDSMGDEDDVDGEBEQSQSEEEYEVEQDDDDSQSEBEVSLPRGRPOVRLPVKTRGKL 1396
: : ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1809 -----GKVVTLTGDE-----DEB----- 1822
: : ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1397 SSSFSRQQQEPGRYPSSQSTPKTTVSSKTKGRSLRKINSAPPTETKSILRIASSTRH 1456
: : ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1823 -----SASTSSSL----- 1830
: : ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1457 SHGPLQADVFFVLLSPRRKRRGRKSANNTPENSFPNFRVIATKSGESQSRVNIASKLS 1516
: : ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1831 -----KEGND-----LQKRKMEINTSIN---LS 1851
: : ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1517 LOESKRKRCKRQSPSPSVTLGRSSGROGGVHELSAFQOLVVELVRHDDSWPFLKLV 1576
: : ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1852 QKESFTSVKPKRDSK-----DLALCSMILTEMETHEDAWPFLIPV 1893
: : ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1577 SKIOVPDYDIIKPIALNIIREKVNKCEYKLASEFIDDIEMFSNCFEYNPRNTSEAKA 1636
: : ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1894 NLKLVPGYKVIKPMDFSTIREKLSGQYPNLETFALDVLRFVDFNCETFNEDDSDIGRA 1953
: : ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1637 GTRLQAFF 1644
: : ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1954 GHNRKYP 1961
: : ||| : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 15
; US-09-839-479-21
; Sequence 21, Application US/09839479
; Patent No. 6727222
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042002
; CURRENT APPLICATION NUMBER: US/09/839,479
; CURRENT FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 21
; LENGTH: 1972
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-839-479-21

Query Match 8.4%; Score 726.5; DB 4; Length 1972;
Best Local Similarity 21.4%; Pred.No.3.6e-44;
Matches 335; Conservative 195; Mismatches 425; Indels 613; Gaps 53;
QY 420 SPANRRGRPP-----KRIHISQEDNVA--- 442
: : ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
```

```
Db 664 SRMRRRGRPPNVGNABFLDNADAKLRLKLOAQEIAQAAQIKLLRLKLOKQEQARVAKEA 723
: : ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 443 --NKOTLASYRSKATKERDKLLKQEE-----MKSLAFE-----KAKLKEKAKADA- 484
: : ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 724 KQQAINAAEKRKQKEQIKIMKQOEKIKRIQIRMEKELRAQOILEAKKKKKEAANAK 783
: : ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 485 -LEAKK--KEKE-----DKEKREELKVIIEERLKKKEKELKLVREKE-- 527
: : ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 784 LLEAEKRIKEKEMRRQAVLLKHQERERRRQHMMKAMEARKKAEKERLQKQKREKR 843
: : ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 528 --REKLBEERKVVYIKQNSKPRDNMECDLDELPE-PTPVKTRLPPEIFGDALMVLEP 584
: : ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 844 LNRKLEQLEMAKELKPNEDMCADQKPLPELPRIPGLVLSGSTFSDCLMVYQF 903
: : ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 585 LNAFGEI--FDLQDEPDPGVTLEVLBEALVG-NDSEGPLCELLFFFTALFOAIAEBEEE 641
: : ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 904 LRNPKVGLGFDVNIQVFN--LSVLQEGLLNIGDSMGVEQDILLVRLISA---AVCDPLGI 957
: : ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 642 VAKEQLTDADTKGCSLKSLLDSCITLSEILRLHILASGADVTSANAKYRYQKRGGFATD 701
: : ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 958 TGYKAKT--ALGHEHLLNVGNRDNVSEILQIF----- 987
: : ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 702 DACMELKLSNPVLVKLSSTSVYDLTPCEKMKIILHALCGKLL--TLVSTRDFTEDYDI 758
: : ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 988 ---MEAHCGQOTELTESLTKTAKFAQHAQTPAKASVLAFILINELACSKSVSEIDKNIDYMSN 1044
: : ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 759 LRQAKQEFRELKABQHRKEREAAAARIRKRKEEKLKEQEKMEKQEKLEKDEEORNSTAD 818
: : ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1045 LRRDKWVV-----EGKLR-----KURITHAKTKGRDTSGG 1075
: : ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 819 ISIGEEB-----REDFDTSIESKDTQEKLQDQDMFTEDDDFGSHKRGRGK 865
: : ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1076 IDLGEQHPLGTPPPGRKRRKGGSDYDDDDDDSDQDQDEDEED---KEDQKGK 1131
: : ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 866 RGQNGFKETFRQEQINCVTRELLTADDEB--ALKQEHQKREKELLEKIOS-----A 914
: : ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1132 KTD-----ICEDEDEGDAQASVEELEKQIEKLSKQSQYRKLFLDA 1172
: : ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 915 IACTNIFPLGRDRMYRYWIFPSIPGLFIE--EDYSGLTEDMLLPRFSSQNNVQSDPQ 972
: : ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1173 SHSLRSVWFGPDRYRRYWIPLRCGGIFVGEESGEGLEELAKEREKLLKKAESVQIKHEM 1232
: : ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 973 VSTKTGEPLMSESTSNIDQG-----VHK-----PNRWCTYSSCEQ 1001
: : ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1233 FET--SGDSLNCSTNDHCEQEDLKEKDNNTLFLQKPGSFSKLSKLELVAXMPPSEVWTP 1291
: : ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1002 KP-----VHK-----PNRWCTYSSCEQ 1018
: : ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1292 KPNAGANGCTLSYQNSGKSLGVSQSTATQSNVEKADSNLNTGSSGPGK--FYSPLPN 1349
: : ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1019 LDQLIEALNSRGHRESAL-----KETLLQEKSRICAQLARFSEKPFHFSDKPOPD--- 1068
: : ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1350 -DQLLKTLTENKRWQFSLLPRTPCDDTSLTHADMSTASLVTTPQSQPPSKSPSPAPLGS 1408
: : ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1069 -----SKPTYSRG-----RS 1078
: : ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1409 SAQNPVGLNPFALSPLOVKGVSMMGLQFCGWPTGVVTSNIPFTLSVPSLGSGLGSEGN 1468
: : ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1079 SNAYDPSQMCABKO-----LEL-RLRDF-----LLDI 1104
: : ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1469 GNSFLTSNVASSKSESVPQNEKATSAQPAAVEAKPVDPPPKPIPEEMQFGWRIIDP 1528
: : ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1105 EDRIYQGTIGAI-----KVTDRIH---IWRSALESGRYELLS--EENKNGI 1145
: : ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1529 ED-----LKALLKVLHLRGIREKALOKQIKHLDYITQACLKNKQVAILNELNENEQV 1582
: : ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1146 IKTV--NEDVEEMIDEQTKVIVKDRLLGIKTKTPTSTVSTWASTPQSVS----- 1192
: : ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1583 TRDIVENWSVEEQAMENDLSVLQOVEDLERRVVASASLQVKGWMCPEPASEREDLVYFEHK 1642
: : ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1193 -----SVVHYL-----AMALFOIEOGIERRFUK--APLDAS 1221
: : ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1643 SFTKCKEHGDEFTGEDESSAHLERKSDNPLDIAVTRLADLERNIERIEDIAP---- 1698
: : ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
```

```
QY 1222 DSGRSYKTVLDRWRBSLSSASQVFLHLSTLDRSVIWSKSIILNARCKICRKGDAENM 1281
Db 1699 -----GLRVRRALSEARGAAQVALLCQLOKSIWEKSIWKYQICRKGDNBEILL 1750
QY 1282 VLCDGCDRGHHYCVRPKLTVPEDGWFCEPRKQ-----RCRLSRQRPSLSEDEDV 1336
Db 1751 LUCDCGDKGCHYCHRPKLTTPDGDWFCPACIAKAGQTLKIKLHVKGKKTNESKK-- 1808
QY 1337 EDSMGEDDEVGDDEEGQSEEEYEVEODEDSQEEBEVSLPKRGRPOVRLPVKTRGKL 1396
Db 1809 -----GKVTLTGDT-----DED----- 1822
QY 1397 SSSFSRGOQPGPYPRSRQOSTPKTVSSKTGRSLRKINSAPPTETKSLRIASRSTRH 1456
Db 1823 -----SASTSSSL----- 1830
QY 1457 SHGPLQADVVELLSPRRKRGRKSANNTPENPNFNRVATIKSSBSRSVNIASKLS 1516
Db 1831 -----KRGND-----LQKRMBSNTSIN-----LS 1851
QY 1517 LOESESRRCKRKRQSPSPVTLGRRSSRGQGVHELSAFEOLVVELVRHDDSWPFLKLV 1576
Db 1852 QOESTSVKPKRDSK-----DLALCSMLITEMETHEDAWFLLPV 1893
QY 1577 SKIQVPDYDIKPIALNIIREKVNCYKILASBFIDDIELMFNSCFEYINPRNTSEAKA 1636
Db 1894 NLKLVGYKVKIKPMDPTIREKSSGQYPNLETFALDVLVDFDNCFTFNEDDSDIGRA 1953
QY 1637 GTRLOAFP 1644
Db 1954 GNMRYKYF 1961
```

RESULT 16

```
US-09-418-710-72
; Sequence 72, Application US/09418710
; Patent No. 6596482
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042001
; CURRENT APPLICATION NUMBER: US/09/418,710
; CURRENT FILING DATE: 1999-10-15
; PRIORITY APPLICATION NUMBER: PCT/JP98/01783
; PRIORITY FILING DATE: 1998-04-17
; PRIORITY APPLICATION NUMBER: JP 9/310027
; PRIORITY FILING DATE: 1997-10-24
; PRIORITY APPLICATION NUMBER: JP 9/116570
; PRIORITY FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 72
; LENGTH: 1969
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-418-710-72
```

Query Match 8.2%; Score 704; DB 4; Length 1969;

Best Local Similarity 21.1%; Pred. No. 1.7e-42;

Matches 331; Conservative 196; Mismatches 426; Indels 616; Gaps 53;

```
QY 420 SPANRRRRGRPP-----KRIHISOEDNVA----- 442
Db 662 SMRRRKGRPPNVGNABFLDNADAKLLRKLQAEARQAQIKLLRKLQKQEQARVAKEAK 721
QY 443 -NKOTLASYSRKATKREKLLKQEE-----MKSLAFE-----KAKLREKADA-- 484
Db 722 KQQAIMAAEKKRQEQIKIHQKQEKIKRIQIIRMKELELRAQOILAEAKKKKEBAANAKL 781
QY 485 LEAKKKEKE-----DKEKREELKKIIVEERLKKKEERLKVREKE----- 527
Db 782 LEAEKRIKEREMRROQAVLLKQERERRRQHMMLMKAMEARKKAEKREKRLKQERDEKRL 841
```

```
QY 528 -REKLREKRYKVEYLKQMSXPREDMECDLKEPE-PTPVKTRLPPEIFGDALMVLBEFL 585
Db 842 NKERKLEQRLEMAKELKKNEDMCLADQKPLPELPRIPLGLVSGSTFSCLMVAQFL 901
QY 586 NAFGEL--FDLQDEPPDGVTVLEVEALVGN--DSEGPLCELLFFFLTAIQAIAESEE 641
Db 902 RNFGKVLGFDVNI DVPN--LSVLQEGILLNIGSMGEVQDILLVRLLSA--AVCDPGLI 955
QY 642 VAKEQLTDADTKGCSLSDLDSDCTLSBILRLHLILASGADVTSANAKYRQKRGGF DATD 701
Db 956 TGYAKT---ALGHEILLNVGNRDVNSILQIF----- 985
QY 702 DACMELRLSNPSLVKKSSTSVYDLTPGEKMKILHAL----CGKLLTLVSTRDFIEDYVD 757
Db 986 --NEAHCGQTELSLTKAFQAHTPAQKAVLAFLINELACSK--SVVSEIDKKNIDYMS 1040
QY 758 ILROAKQEFRELKAEQHRKEREAAARIRKKEKLEKEQEQMKKEKQELKXEDGORNSTA 817
Db 1041 NLRDRKNV-----EGKLR-----KLRIITHAKTKGRDTSG 1071
QY 818 DISIGEEB-----REDFDTSIESKDEQKELDODMFTEDDDPGSHKRGRRG 864
Db 1072 GIDGEEQHPLGTPTPGRRRRKGGDSYDDDDDDSDQDDEDEDEED---KEDQKG 1127
QY 865 KRQONGKFEFTROEQINCVTRELLTADREE--ALKQEHORKEKELLEKIQS----- 913
Db 1128 KKTD-----ICEDDEGDAQSAVEELEKQIEKLSKQSQSYRRKLFD 1168
QY 914 AIACNTIPPLGRDMRYRYWIFPSIPGLFIE--EDYSGLTEDMLLPRESSFQNNVQSODP 971
Db 1169 ASHSLRSVMFGPDYRRYRYYWILPRCGGIFVEGMESEGLEIEAKEREKLLKKAESVQIKEE 1228
QY 972 QVSTKTGPELMSESTSNIDQG-----PRDHSVOL 1000
Db 1229 MFET--SGDSLNCSTNDHCEQKEDLKEKDNVTLFQKPGSFSLKSLLEVAKMPPESEVMT 1287
QY 1001 PKP-----VHK-----PNRWCIFYSCCE 1017
Db 1288 PKPNAGANGCTLSYQNSGKHSLSGVSQSTATQSNVEKADSNLNTGSSGPGK--FYSPLP 1345
QY 1018 QLDOLIEALNSRGHRESAL-----KETLLQEKSRICAOARFSEKHFHSDKPOPD--- 1068
Db 1346 N-DQLLKTLEKRCQWFSLLPRTPCDDTSLTHADMSTASLVTPQSQQPSKSPSPFPAPLG 1404
QY 1069 -----SKPTYSRG-----R 1077
Db 1405 SSAQNPVGLNPFALSPLOKGVGVMGLQFCGWPTGVVTSNIPPTLSVPSLGSGLGSEG 1464
QY 1078 SSNAYDPSOMCAEQ-----LEL-RLRDF-----LLD 1103
Db 1465 NGNSFLTSNVASSKSESPPVQNEKATSAQPAAVEVAKPVDPPSPKPIPEEMQFGWRIID 1524
QY 1104 IEDRIYQGTLCAL-----KVTDRH---IWRSALESGRVELLS-ENKENG 1144
Db 1525 PED-----LKALLKVLHRLGIREKALQKQKHLDYITQACLKNKDVAILIENENENQ 1578
QY 1145 IIKTV--NEDVEEMEIDQTKVIVKDRLLGKTTETPTSTVSTNASTPQSVS----- 1192
Db 1579 VTRDIVNWSVEEQAMENDLSVLOQVEDLERRVASASLVQKGMWCPPEPASEREDLVYFEH 1638
QY 1193 -----SVVHYL-----AMALFQIEQGIERRERFLK--APLDA 1220
Db 1639 KSFTKLCHEHDGEFTGEDESSAHALERKSDNPLDIATVTRLADLERNIERIEEDAP--- 1695
QY 1221 SDSGRSYKTVLDRWRBSLSSASQVFLHLSTLDRSVIWSKSIILNARCKICRKGDAEN 1280
Db 1696 -----GLRVRRALSEARGAAQVALLCQLOKSIWEKSIWKYQICRKGDNBEEL 1746
QY 1281 MVLCDGCDRGHHYCVRPKLTVPEDGWFCEPRKQ-----RCRLSRQRPSLSEDED 1335
Db 1747 LLLCGCDKGCHYCHRPKLTTPDGDWFCPACIAKAGQTLKIKLHVKGKKTNESKK- 1805
```

```
QY 1336 VEDSMGDEVDGDEEQSEEEYVEVEDDSDQEEBESVLPKRGRPOVRLPVKTRGK 1395
      : : : : :
Db 1806 -----GKVTLTGDT-----DED----- 1819
QY 1396 LSSSFSSRGQOQEPGRYPSRQSQTPKTVTSKGTGSLRKINSAPPTKSLRIASRSTR 1455
      : : : : :
Db 1820 -----SASTSSL----- 1827
QY 1456 HSHGFLQADVVELLSPRKRGRKSANNTTPNSFNFRFVIATKSEQSRSVNIASKL 1515
      : : : : :
Db 1828 -----KGNKD-----LQKKEEINTSIN-----L 1847
QY 1516 SLQSESEKRRCKRQSPSPVTLGRSSGROGVHLSAFEQLVVELVVRHDDSWPFLKL 1575
      : : : : :
Db 1848 SKQESFTSVKPKRDSK-----DLALCSMILITETHEDAWPFLP 1889
QY 1576 VSKIOVPDYDIIKPIALNIIRKVNKEVYKLASEFTDDIELMPSNCFNPNRNTSAK 1635
      : : : : :
Db 1890 VNLKLVPGYKVKIKPMDFSTIREKLSGQYPNLETFALDVRLVFDNCETFEWDDSDIGR 1949
QY 1636 AGTRLQAFF 1644
      : : : : :
Db 1950 AGHNURKYF 1958

RESULT 17
US-09-839-479-71
; Sequence 71, Application US/09839479
; Patent No. 6727222
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042002
; CURRENT APPLICATION NUMBER: US/09/839,479
; PRIOR FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 71
; LENGTH: 1969
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-839-479-71
```

```
Query Match      8.2%; Score 704; DB 4; Length 1969;
Best Local Similarity 21.1%; Pred. No. 1.7e-42;
Matches 331; Conservative 196; Mismatches 426; Indels 616; Gaps 53;

QY 420 SPANRRRCRPP-----KRIHSQEDNVA----- 442
      : : : : :
Db 662 SMRRKGRPNVNGNAEFLDNADAKLRKLAQEARQAQIKLRKQKQEQARVAKEAK 721
QY 443 -NKQTLASYRSKATKERDKLLKQEE---MKSLAFE-----KAKLRKREKADA-- 484
      : : : : :
Db 722 KQQAIMAAEERKQKEQIKIHQKEKIKRIQIRMEKELRAQQILEAKKKKEEANAACL 781
QY 485 LEAKKKEKE-----DKEKGREBKXIVBEERLUKKEEKEKRLKVERKE--- 527
      : : : : :
Db 782 LEAEKRIKERMRROQAALLKQERERRRRQHMLMKAMEARKKAEEKERLQEKRDKRL 841
QY 528 -REKLEREKRYVEYLQWSPREDMECDLKEKPE-PTPVKTRLPPEIFGDALMWLEFL 585
      : : : : :
Db 842 NKERLEORRLEMAKELKPKNEDMCIADQKPELPRIPLGLVLSGSTFSDCLMWVQFL 901
QY 586 NAFGEL--FDLQDFPPDQVTLLEVEEALVGN--DSEGPLCELLFFFLTAIFAQAEABEE 641
      : : : : :
```

```
Db 902 RNFGKVLGFDVNIIDVPN---LSVLQEGILLNIGDSMGVEQDLLVRLLSA-----AVCDPGLI 955
QY 642 VAKQLTADTKGCSLAKSLDLSDCTLSFILRLHLTLASGADVTSANAKRYQKRGGFDAITD 701
      : : : : :
Db 956 TGYKAKT---ALGBHLLNVGNVRDNVSEILQIF----- 985
QY 702 DACMELRLSNPSLVKLSSTSVYDLTPGEKMKILHAL---CGKLLTLVSTRDIEDYVD 757
      : : : : :
Db 986 ---NEAHCGGTETLSLTKAFOAHTPAQKAVLAFLINELACSK--SVSEIDKNIDYMS 1040
QY 758 ILROAQEFRELKAEQHRKEREAAAIRKKEKPELKEQEKMEKEKLEKEDQORNSTA 817
      : : : : :
Db 1041 NLRDKNVV-----EGKLR-----KLRIHAKKTKGRDTSG 1071
QY 818 DISIGEE-----REDFPDISBKOTEQEKELDQDMFTEDEDDPGSHKGRGRG 864
      : : : : :
Db 1072 GIDLGEBOHPLGTTPGKRKRKGGSDYDDDDDDSDQGDDEDEED---KEDQKG 1127
QY 865 KRQONGKFEFRQBCQINCVTRRELITADEEE---ALKQEHQKKEKELLEKIQS----- 913
      : : : : :
Db 1128 KXTD-----ICEDDEGDQAASVEELEKQIEKLSKQSQSYRKLFD 1168
QY 914 AIACNTIPLGRDMRYRMYIPFISIPGLFIE--EDYSGLTEDMLLPPRSPFQNNVQSDP 971
      : : : : :
Db 1169 ASHSLRSVMFGPDYRRRYRWILPRCGGIFVEGMEGEGLEEIAKEREKLLKKAESVQIKEE 1228
QY 972 QVSTKTGEPLMSESTSNIDQG-----PRDHSVQL 1000
      : : : : :
Db 1229 MFET-SGDSLNCNNTDCEQKEDLKEKDNTLNLFQKPGSFSLKJLLEVAOMPESEVMT 1287
QY 1001 PKP-----VHK-----PNRWCFYSCE 1017
      : : : : :
Db 1288 PKPNAGANGCTLSYONGSKHSLGSVQSTATQSNVEKADSNLNTFGTSGSGGCK--FYSLPL 1345
QY 1018 QLDQLEBALNSRGHRESAL-----KETLQEKSRICAQALAREEKFHFSDKPPD--- 1068
      : : : : :
Db 1346 N-DQLLXTLTKENRQWFSLLPRTPCDDTSLTHADMSTASLVTPQSQPSPSPSTPAPLG 1404
QY 1069 -----SKPTYSRG-----R 1077
      : : : : :
Db 1405 SSAQNPVGLNPFALSPLOVKGVSMMGLQFCGWPTGVVTSNIPTLSVPSLGSGLGSEG 1464
QY 1078 SSNAYDPSOMCAEQ-----LEL-RLRDF-----LLD 1103
      : : : : :
Db 1465 NGNSFLTSNVASSKSESPVQNEKATSAQAAVEVAKVPDPFPPKPIPEMQFQFWWIID 1524
QY 1104 IEDRIYOGTILGAI-----KYTDRH---IMRSALBSGRYELLS--EENKENG 1144
      : : : : :
Db 1525 PED-----LKALLKVLHLRGIREKALQKQIKQHLDYITQACLKNKQDAIIELENEENQ 1578
QY 1145 IIKTV--NEDVEEIDEQTKVIVKORLLGIKTKETPTSTVSTNASTPOSVS----- 1192
      : : : : :
Db 1579 VTRDIVENWSVEEQMEMDLSVLQOVEDLERRVASASLQVKGWMCPEPASEREDLVYFEH 1638
QY 1193 -----SVVHYL-----AMALFQIBOQIERFLK--APLDA 1220
      : : : : :
Db 1639 KSFTYKCKEHDGFTGEDESAAHALERKSNPNLDIAVTRLADLERNIERIEADIAP--- 1695
QY 1221 SDSGRSYKTVLDRWRSELLSSASLSQVFLHSLTLDRSVIWSKILNARCKICRKKGAEN 1280
      : : : : :
Db 1696 -----GLRVWRRLSEARSAAQVALCIIQLOKSIAMEKSIIMKYVCQICRKGDNEL 1746
QY 1281 MVLCDGCDRHHYCYVRPKLTVPEGDWFCPECRPKO-----RCRRLSFRQRPSELESD 1335
      : : : : :
Db 1747 LLLCDGCDKCHTYCHRPKITTIPDGDWFCPACIAKASGQTLKIKLHVKGKKTNESKK- 1805
QY 1336 VEDSMGDEVDGDEEQSEEEYVEVEDDSDQEEBESVLPKRGRPOVRLPVKTRGK 1395
      : : : : :
Db 1806 -----GKVTLTGDT-----DED----- 1819
QY 1396 LSSSFSSRGQOQEPGRYPSRQSQTPKTVTSKGTGSLRKINSAPPTKSLRIASRSTR 1455
      : : : : :
Db 1820 -----SASTSSL----- 1827
```



```
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042001
; CURRENT APPLICATION NUMBER: US/09/418,710
; CURRENT FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 1878
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(1878)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-09-418-710-13

Query Match          7.8%; Score 671.5; DB 4; Length 1878;
Best Local Similarity 19.8%; Pred No. 3.9e-40;
Matches 305; Conservative 197; Mismatches 399; Indels 639; Gaps 44;

QY 424 RRRGRPPK-RIH--ISQEDN-----VANKOTLASYSKATKRDKLLKQBEEMKSL 470
DB 644 RGRGRPPVKITELANKTNRPLKXLEAQETLNEEDKAKIAKSKKMKQKVRQGBCLTTI 703
QY 471 AFEKAKLKREKADAL---EAKKEKEDKEK---KREELKKIVVEE-----RLKKKEE--- 516
DB 704 QGARINKRKQETKSLKHKEAKKSAEKEGKTKOEKLEKVKREKKEKVKMKKEEVTK 763
QY 517 -KERLKVREKERELKREKRYVEYLQWKSPPREDECDLDELPEPTPVK-TFLPPEI 574
DB 764 AKPACKADKTATLORLEQRQKQMLLEMKKPTEDMCLTDHQPDPFSRVPLGLTLPSPA 823
QY 575 FGDALMLVLEFLNAFQEL--FDLQDFPDGVTLEVLVEALV--GNDSEGPLCELLFFFLTAI 631
DB 824 FSDCLTIVFELHLSFKVLGFDPAKQVP---SIGVLQELGLCGQSLGVEVDLLVRLKAA 880
QY 632 F-----QAIAEBEEBAKEQLTDADTKGCSLSKSLDLSCTISEILRLHIILASGADV 682
DB 881 LHDPGPSYCSQSLKILGKVSSEIPLTRDN-----VSEILRCFLMAYGV-- 923
QY 683 TSANAKRYKRGGFDDATDACCMELRNLNPSLVKLSSTSVYDLTPGEMKIL-----HAL 738
DB 924 -----XPALCDRLRTQFPQAPQPOQKAALVAFPPVHEL 955
QY 739 CGKLLTLVTRDFIEDYVDILRQAKQEPRELKAEQHERBEAARIRKRKEELKQEQ 798
DB 956 NGSTL-----IINEIDKTLSEMS--YRKNKIVGELRLKTLVLAK----- 995
QY 799 KMKKEQKLEDEQRNSTADISIGEEEREDFTSTIESKDTQKELDQDMFTDEDDPGSH 858
DB 996 -----RTGRSEVEMGRPE---ECLGRRSSRIMETSGMEEEEEESTAA 1037
QY 859 KRGGRKEG-----QNGFKETRQEQINCVTRELLTADBEALKQEHQREKELLEKIQ 912
DB 1038 VPGRRGRDGEVDATASSIPELQ-----TEKLSKROLFRKKLLHSSQ 1082
QY 913 SAIACTNIFPLGRDMRYRYIFPISPLGLFTEEDYSGLTDEMMLLP----- 957
DB 1083 MLRAVS-----LQODRYRRYVLPVLAGIFVEG-----TEGNLVPVEVKKETSLKVA 1133
QY 958 -----RPSSFQNNVQSD----- 970
DB 1134 HASLNPALFSKMELAGSNNTTASSPARARSRLKTKPGFMQPRHFKSPVQGDSEPOQAQ 1193
QY 971 -----PQV-----STKTGE 979
DB 1194 LQPEAQLHVPAPQPOLQLQSHKGFLEQEGSPLSLQSQSHLSQSASFSLWSLQTSQHS 1253
```

```
QY 980 PLMSESTSNIDQGR-----DHSVQLP----- 1001
DB 1254 SLLSSSVLTPDSFGKLDLPAPSQPPEPEDEABSSPDLOAFWFNISAQMPCNAAPTPL 1313
QY 1002 -----KPVHKPNR----- 1009
DB 1314 AVSEBQPTPSPQQLASSKPMNRPSAANPCSPVQFSSTPLAGLAPKRRAGDPCGMPQSP 1373
QY 1010 -----WCFYSSCEQLDQIETALNSRGH 1031
DB 1374 LGQPKRCRPPSKFKQMEQRYLTQLTAQVPPPMCSGMMWIPDPEMLDALMKALHPRGI 1433
QY 1032 RESALKETLLQEKSRICAQLARFSEKPHFSDKQPDSPKPYSGRGRSNAYDPSQCMCAEK 1091
DB 1434 REKALHKLHNKRDPLQEVCLRPSADPIF-----EPROLPAFQEGIMS--WSPKEKTYE- 1485
QY 1092 QLELRRLRDFLLDIEDRIYQGTLGAIKVTDRHIMWSALESGRYELLSEENKENGIIKTVNE 1151
DB 1486 -TDLAVLQWVELEQRVIM-----SDLQIRGWTCPSPD-----STRE 1521
QY 1152 DVEEMEIDEQTKVIVKDRLLGIKTETPTSTVNASTPQSVSSVHYLAMALFQIRQGIER 1211
DB 1522 DLAYCEHLSDSQEDITWRGPGREGLAPOKRTNP-----LDLAVRLAALQONVKR 1572
QY 1212 RFLKAPLDASDGRSYKTVLDR-----WRESLSSASL 1244
DB 1573 RYLRPL-----WPTHEVLEKALLSTPNGAPEGTTTTEISYEITPRIRWQTQLQCRSA 1627
QY 1245 SQVFLHLSTLDRSVIWSKSIILNARCKICRKKGDANWVLCDGCRGHHTYCVRPKLKTVP 1304
DB 1628 AHVCLCLGHLERSIAWEKSVNKNVTCLVCRKGDNDDEFLLDCGDCRGCHYCHRPKQEA 1687
QY 1305 EGDWFCPECRKQCRRLSRFRQPSLESDEDEVEDSMGDEDEVDDEEGQSEEEVEVE 1364
DB 1688 EGDWFCVCLAQ-----VEGEFTQKQFPKRGQKRSQSYSLN 1725
QY 1365 QEDDSQEEEBEVSLEKRGPRQVRLPVKTRGKLSSSFSSRQGOQBPGRYPSPRSQSTPKTT 1424
DB 1726 FSEG-----GR-----RRVLLKGR-----ESPAAGPRYSE----- 1752
QY 1425 VSSKTGRSLRKINSAPPTETKSLRIASRSTRHSHGFLQADVVELLSPRRKRGRKKSANN 1484
DB 1753 -----ERLSPSKRR----- 1762
QY 1485 TPENSPFNPRVIATKSSEQSRSVNIASKLSQESKRCRKRQSPSPSVTLGRSS 1544
DB 1763 -----LSMRNHS----- 1770
QY 1545 GROGGVHELSAFEQVLVVELVRHDSWPFLKLVSKIQVDPYDYDIKKPTALMIIREKVNC 1604
DB 1771 -----DLTFCFILLMEMESHDAWPPXEPVNPRLVSGYRRIIKVPMDFSTMRELLRG 1823
QY 1605 EYKLASFIDDIELMFNCFEYNPNRNTSEAKAGTRLQAFF 1644
DB 1824 GYTSSEFAADALLVFNQCTFNEDDSEVGKAGHIMRRFF 1863
```

RESULT 21

```
US-09-839-479-13
; Sequence 13, Application us/09839479
; Patent No. 6727222
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042002
; CURRENT APPLICATION NUMBER: US/09/839,479
; CURRENT FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
```


; OTHER INFORMATION: leic acid sequence of AAP- US-09-764-176-7									
Query Match	4.2%;	Score	361.5;	DB 4;	Length	1400;			
Best Local Similarity	18.8%;	Pred. No.	1.9e-17;						
Matches	265;	Conservative	184;	Mismatches	457;	Indels	505;	Gaps	49;
Qy	436	SOEDNVANKQT	LA	YRSKAT	KERDKLLKQ	ENKSLAF	EAKAKLRPK	-----ADAL	485
Db	177	SQODNSSRES	--PSLEDBET	KKEETPKQ	EQK-----E	SEKWKSE	QEQMDLEN	STANVL	230
Qy	486	E--AKKKEK	DEK-----K	RELKVI	VEERLKK	EERLKV	-----	523	
Db	231	BETTVKKE	DEKELV	LVIVL	KEPLPEN	EKKIIE	ESDSFK	IKVEK	290
Qy	524	-----REK	ER-----	KLRE	KKVY	YLKQ	MSKP	548	
Db	291	DPKDTK	SMXEP	VAGEP	ERIBFG	GNIK	SSHEIT	ETEST	350
Qy	549	REDME	-----	CDL	KELP	ETPV	KTRLP	PFIFG	601
Db	351	SDDFSP	VKGLCK	SV-----	TPTKE	FLKDE	IKQEE	-----ETCK	400
Qy	602	VTLEVL	EALV	GNDS	EGPL	CELL	PFPLTA	IFOA	661
Db	401	VNGEVS	DVAP	NFKTE	PI--ETK	FY-----	ETKE	ESYSP	450
Qy	662	--LDSC	TLSE	ILR--L	HILAS	GADV	TSA	NAKY	718
Db	451	SVITSM	KTGE	LEKET	APLRK	AD--SS	IVLH	SQAK	509
Qy	719	SSTS	YVDL	TPG	KMKIL	HALC	KLTL	VSTR	775
Db	510	SSKTAL	SSTES	CTMK-----	GEES	PKTK	DXKRP	PILE	563
Qy	776	KERBA	AAIR	KRKE	EKKE	KBOE	QKQ	KEQ-----EK	823
Db	564	PIPE	-----	VPK	STLE	SEK	PGS	PEAA	619
Qy	824	EERED	FTS	IES	KOT	EQ	KELD--	QDM	881
Db	620	VKKV	LETL	KED	SET	KVEM	DLNA	QTS	668
Qy	882	CVTREL	LTA	DEE	ALQ	BOR	KE	LEBK	941
Db	669	LVPE	BETT	ASE	NT	IT	SRQ	EGIK	710
Qy	942	FIEED	YSL	G	LT	EDML	PRFSS	QNNV	1001
Db	711	VLE	-----	PEN	KQ	ET	KEE	K--TNV	741
Qy	1002	KPVH	KPNR	WC	FV	SSCE	QDL	LI	1061
Db	742	RPTAK	-----	VAE	IRD	QK	ADK	KRG	781
Qy	1062	SDKP	QD	SK	PT	YS	RGR	SSN	1121
Db	782	LKKSE	KTNS	KVSK-----					
Qy	1122	HIW	SAL	ES	G	RY	ELL	S--BEN	1180
Db	802	VRWT	G	SR	TR	G	R	WK	849
Qy	1181	VSTNA	ST	Q	S	V	S	V	1240
Db	850	-----							
Qy	1241	SASL	SQ	VL	HL	ST	LD	R	1300
Db	850	-----							
Qy	1301	KTV	PE	G	D	W	F	C	1343

885

MIIPGEWFCPPQO

QKLLCEKLEBQ

LDLVALKKK

ERARERK

LVTVVGIS

ENIIP

PPQ

944

1344

DDEVDGDEEQ

SEEEYEV-----

EQEDDSQ

1372

945

EPDFSEDOE

KKKSKANLL

ERRSTR

TRKCI

SYR

DFDE

FAIDE

IEDD

IK

EADGG

1004

1373

EEVSLPK

GRQV

RLPV

TK

RG

KL

SS

SSFS

1401

1005

VGRGDISTIT

GRGKDIST

ILDEER

KENK

RPQ

AAAA

ARRK

RRRL

NDL

SD

SNL

DEES

1064

1402

SRGQOE

PGYPS

R

SO

Q

TP

KT

VS

SK--

TGR

SL

R

K

I

N

S

A

P

T

E

T

K

S

L

R

I

A

1450

1065

EDEFKIS

Q

DE

F

V

S

D

N

E

S

E

D

P

P

S

N

D

S

O

T

F

C

S

R

R

L

R

H

P

R

P

Q

R

R

L

R

K

1124

1451

SRSTRH

SGP

LIQ-----

ADVF

ELL

SP

RR

K

R

G

R

K

S

A

N

T

P

E

N

S

N

F

1492

1125

TPKKY

SD

DEE

SE

NS

R

DS

ES

DF

SD

F

VE--

TR

R

R

S

R

N

Q

R

Q

N

Y

K

E

D

S

E

D

1183

1493

PNFR

V

I

A

T

K

S

E

Q

S

R

N

I

A

S

K--

LS

Q

E

S-----

ES

K

R

C

R

K

1529

1184

G-----

SOK

S

L

R

G

K

E

I

R

R

V

H

K

R

L

S

S

E

S

E

S

Y

L

S

K

N

S

E

D

E

L

A

K

S

K

R

S

V

R

K

R

G

R

S

T

1238

1530

Q

S

P

S

P

T

L

C

R

R

S

G

Q

G

V

H

E

L

S

A

F

Q

L

V

E

L

V

R

H

D

S

M

P

F

L

K

V

S

K

I

O

V

P

D

Y

D

1586

1239

EY

S

E

A

D

E

E

E

E

E

K

P

S

R

K

R--

L

H

R

I

E

T

D

E

E

S

C

D

N

A

G

D

A

N

Q

P

R

D

S

Q

P

R

V--

L

P

S

E

Q

1294

1587

I

I

K

P

I

A

L

N

I

R

E

K

V

N

K

C

E

V

K

L

A

S

E

F

I

D

D

I

E

1617

1295

ST

K

P-----

Y

R

I

E

S

D

E

E

D

F

E

1311

RESULT

23

US-09-418-710-38

Sequence 38, Application US/09418710

Patent No. 6596482

GENERAL INFORMATION:

APPLICANT: Jones, Michael H.

TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR

FILE REFERENCE: 06501-042001

CURRENT APPLICATION NUMBER: US/09/418,710

CURRENT FILING DATE: 1999-10-15

PRIOR APPLICATION NUMBER: PCT/JP98/01783

PRIOR FILING DATE: 1998-04-17

PRIOR APPLICATION NUMBER: JP 9/310027

PRIOR FILING DATE: 1997-10-24

PRIOR APPLICATION NUMBER: JP 9/116570

PRIOR FILING DATE: 1997-04-18

NUMBER OF SEQ ID NOS: 73

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 38

LENGTH: 59

TYPE: PRT

ORGANISM: Homo sapiens

US-09-418-710-38

Query Match 3.6%; Score 313; DB 4; Length 59;

Best Local Similarity 100.0%; Pred. No. 6.2e-16;

Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1569

SWPFLK

LVSK

IQV

PDY

YD

II

K

P

I

A

L

N

I

R

E

K

V

N

K

C

E

V

K

L

A

S

E

F

I

D

D

I

E

L

M

F

S

N

C

F

E

Y

N

1627

Db

1

SWPFLK

LVSK

IQV

PDY

YD

II

K

P

I

A

L

N

I

R

E

K

V

N

K

C

E

V

K

L

A

S

E

F

I

D

D

I

E

L

M

F

S

N

C

F

E

Y

N

59

RESULT

24

US-09-418-710-50

Sequence 50, Application US/09418710

Patent No. 6596482

GENERAL INFORMATION:

APPLICANT: Jones, Michael H.

TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR

FILE REFERENCE: 06501-042001

CURRENT APPLICATION NUMBER: US/09/418,710

```
; CURRENT FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 50
; LENGTH: 59
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-418-710-50

Query Match          3.6%; Score 313; DB 4; Length 59;
Best Local Similarity 100.0%; Pred. No. 6.2e-16;
Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1569 SWPFLKLVSKIQVDPDYDIKKPTALNIIREKVNKCEYKLASEFIDDIELMFNSCFEYN 1627
Db 1 SWPFLKLVSKIQVDPDYDIKKPTALNIIREKVNKCEYKLASEFIDDIELMFNSCFEYN 59

RESULT 25
US-09-839-479-37
; Sequence 37, Application US/09839479
; Patent No. 6727222
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042002
; CURRENT APPLICATION NUMBER: US/09/839,479
; CURRENT FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 37
; LENGTH: 59
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-839-479-37

Query Match          3.6%; Score 313; DB 4; Length 59;
Best Local Similarity 100.0%; Pred. No. 6.2e-16;
Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1569 SWPFLKLVSKIQVDPDYDIKKPTALNIIREKVNKCEYKLASEFIDDIELMFNSCFEYN 1627
Db 1 SWPFLKLVSKIQVDPDYDIKKPTALNIIREKVNKCEYKLASEFIDDIELMFNSCFEYN 59

RESULT 26
US-09-839-479-49
; Sequence 49, Application US/09839479
; Patent No. 6727222
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042002
; CURRENT APPLICATION NUMBER: US/09/839,479
; CURRENT FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
```

```
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 49
; LENGTH: 59
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-839-479-49

Query Match          3.6%; Score 313; DB 4; Length 59;
Best Local Similarity 100.0%; Pred. No. 6.2e-16;
Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1569 SWPFLKLVSKIQVDPDYDIKKPTALNIIREKVNKCEYKLASEFIDDIELMFNSCFEYN 1627
Db 1 SWPFLKLVSKIQVDPDYDIKKPTALNIIREKVNKCEYKLASEFIDDIELMFNSCFEYN 59

RESULT 27
US-08-056-200-94
; Sequence 94, Application US/08056200
; Patent No. 5616500
; GENERAL INFORMATION:
; APPLICANT: Steinert, Peter M.
; APPLICANT: Lee, Seung-Chul
; APPLICANT: Kim, In-Gyu
; APPLICANT: Chung, Soo-Il
; APPLICANT: Park, Sang-Chul
; TITLE OF INVENTION: Trichohyalin and Transglutaminase-3 and
; TITLE OF INVENTION: Methods of Using Same
; NUMBER OF SEQUENCES: 117
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 620 Newport Center Drive, Sixteenth Floor
; CITY: Newport Beach
; STATE: CA
; COUNTRY: U.S.A.
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/056,200
; FILING DATE: 30-APR-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Fedrick, Michael F.
; REGISTRATION NUMBER: 36,799
; REFERENCE/DOCKET NUMBER: NIH054.001A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (714) 760-0404
; TELEFAX: (714) 760-9502
; INFORMATION FOR SEQ ID NO: 94:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1898 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; Molecule Type: protein
US-08-056-200-94

Query Match          3.6%; Score 310; DB 1; Length 1898;
Best Local Similarity 20.2%; Pred. No. 1.9e-13;
Matches 283; Conservative 203; Mismatches 459; Indels 454; Gaps 60;

Qy 350 ELHESAIVKATQISRKHFLSRDK-----LKLFLKQHCPEQPGVIKKASLSYKI 401
Db 484 ERHEQ-----ERRKQQLKQDEERRRRRLKLEERREQER----- 521
```


Db 484 ERHEQ-----ERRKOQLKRDQBEERRRWLKLSEERREQOER----- 521

Qy 402 AEQDPSYFFPDDPTTFIFSPANRRGRPPKRIHISQEDNVANKOTLAS---YRSKATKER 458

Db 522 REQQ-----LRREQERREQLKR-----QEEERLQORLSEQQLRQEERL 565

Qy 459 DKLKQOEMKSLAFK---AKLR---EKADALEAKKKEKEDKREKRLKIVVEERLKK 513

Db 566 EQLLKREKKEQLQERREORLKRQEERDQLLKREERROQLKREGEERL--EQRL-K 622

Qy 514 KEERLKVREKEREKUR---BEKKYVYLVKQMSKPREDMCEDDLKELPEPTPVKTR 569

Db 623 REEVERLEQE-ERRDERLKREPEERH-ELLK--SEEQERREHEQLR-EQERREQR 677

Qy 570 LPPIFGDALMWLEFLNAGELFDLQDFPDGVTLEVLSEALVGNDSGELCELLFFFLT 629

Db 678 LKRE-----EEB-----ERLEORLKREHEE-----E 698

Qy 630 AIFOAIEEEVEVAKEQLTDADTKGCSLKSLDLDSCITSEILRLHILASGADVTSANAKY 689

Db 699 REQELAEEOQARERIKSRPKW-----QWQLESEADARQSKVLL 740

Qy 690 RYQKRGDPDADAMELRNLSPVLVKLSYSTSYVDLTPGKMKILHALGKLLTVLSTR 749

Db 741 EAPQAGRAEAPQE--QEKRRRESELOWEERAHROOQOBEQR-----R 783

Qy 750 DF-----IEDYVDILRQ-----AKQEFRELAQEHK-----EREAAARIRKK 789

Db 784 DFTWQWAEKSEGRORLSRAPPLREQRQRLAEERQOERQFLPEEKEQGRQRR 843

Qy 790 E-----EKLKEQKMKKEQKLEKEDQORNSTADISIGEEERDF 829

Db 844 EREKELQLEEEQLORRERAQQLQEEEDGLOEQERRRQORQDKRWQL-BEERKR 902

Qy 830 DTSIESKDEQEL--DQDMFTEDDDPGSHKRRGRKGQNGKPEFTROQINCVTREL 887

Db 903 RHTLYAKPALQELKEQQLQEEELQREERKRRQEQE--RQYREEBQLQOEEBQL 960

Qy 888 LTADREEALKQHQ---RKEKELLEKTQSACTNIPLG-----RDRMYRYWI 934

Db 961 LREERKRRROERQRYVKDKKLOQKEQL-----LGEPEKRRRQREKKYR--- 1008

Qy 935 FPSIPGLFIEDYSGLTDMLLPFPSPQNNVQSDPOVSTK-----TGEPLMSESTSI 989

Db 1009 -----EBEELQOEEBQLL-BEERKRRQEWERYRKKDELQOEEBQLLEERKR 1058

Qy 990 DQGRDHSVQLPKPVKPNRWCF-----YSSCEOLOLIEAL----- 1026

Db 1059 RQEREROYREEBEELQOEEBQLGEERETRRRQELERYRKEEELQOEEBQLLEPEKR 1118

Qy 1027 -NSRGHRESALKETLLOEKSRIC-----AQLAR-FSEE-----KFHFSDKPO- 1066

Db 1119 RQEREROQREEBEELQOEEBQLLEERKRRRQELERYRKEEELQOERQRYRDEQR 1178

Qy 1067 -----PDSK-----PTYSRGRSSNAY---DPSQWCAEQLELRDLPLLOIEDRIY 1109

Db 1179 SDLKWQHEPEKENAVRONKYCKRENEQFLQEDSQV-RDRQSQDQLQHLGQOER-- 1235

Qy 1110 QGTIGAIVTDRIHWSA-----LESGRYELLSEENKENGHIKTNEDVE 1154

Db 1236 -----DREQERRWQOANRHFPPEEQLEERQEAKEAKRDRKSQEQK-----LLREERE 1284

Qy 1155 EMEIDEQT--KVIVKDRLLGKITETPTSTVSTNASTPQSVSSVHYLAMALFOIQGLERR 1212

Db 1285 EKRRQETDRFREEBEQLQOREEQP-----LLRQERDRK 1319

Qy 1213 FLKAPLQDASDGRSVKTVLDRWRESLLSSASLSQVFLHLSLTDRSVIWSKILNARCKIC 1272

Db 1320 FREELLHQOGRKFLSEEBQLRE-----ERERKFLKEEQQLRL-- 1359

Qy 1273 RKKGDAENWVLCGDCGRGHHTYCYVRPKLTVPEGDMFCPECRPKORCRRLSFRQRPLES 1332

Db 1360 -----EREQLRQDRDKFREEBEEOQLS 1380

Qy 1333 DEDVEDSMGGEDDDVGDGDEEGEYVEVEDDDSQEEEVSLPKRGRPOVRLPVKT 1392

Db 1381 ROEDRFRREEQOVRROERERKFLSEEQQLRQERHRKFRFEEQQLQEREEQLHRQERD 1440

Qy 1393 RGKLSSFSRSGQQOQEGRYPSRSQOSTPTKTVSSKTGRSLR--KINSAPPTETKSLRIA 1450

Db 1441 RKFLEEBQQLRRQERD-----RKFREQELASQEP-ERKFLEEE 1477

Qy 1451 SRSTRHSHGLQADVFEVLLSPRRKRRGRKRSANNTPENSPNFRVIATKSSSQSRSVN 1510

Db 1478 QQLHROQR---QRKFLQEQQLRRQERGOQRORQDRD-----KFR-----EEBQLRQER 1523

Qy 1511 IASKLSIQESSEKRCR-----KQSPS---PSPVTLGRSSRGOGGVHLSAF---BQLV 1560

Db 1524 BEQLLSQERDKRFLREEQKVRQEQERKFMWEDQQLRROEQQLQERDKKFRFEDSOLL 1583

Qy 1561 VE-----LVRHDDSWPFLK 1574

Db 1584 QEREEQQLHRQERDKFLE 1602

RESULT 29

US-09-538-092-1280

; Sequence 1280, Application US/09538092

; Patent No. 6753314

; GENERAL INFORMATION:

; APPLICANT: Giot, Loic

; APPLICANT: Mansfield, Traci A.

; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same

; FILE REFERENCE: 15966-542

; CURRENT APPLICATION NUMBER: US/09/538,092

; CURRENT FILING DATE: 2000-03-29

; PRIOR APPLICATION NUMBER: 60/127,352

; PRIOR FILING DATE: 1999-04-01

; PRIOR APPLICATION NUMBER: 60/178,965

; PRIOR FILING DATE: 2000-02-01

; NUMBER OF SEQ ID NOS: 1387

; SOFTWARE: CurapatSeqFormat Version 0.9

; SEQ ID NO 1280

; LENGTH: 1898

; TYPE: PRT

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc_feature

; LOCATION: (0)...(0)

; OTHER INFORMATION: Polypeptide Accession Number Q07283

US-09-538-092-1280

Query Match 3.6%; Score 310; DB 4; Length 1898;

Best Local Similarity 20.2%; Pred. No. 1.9e-13;

Matches 283; Conservative 203; Mismatches 459; Indels 454; Gaps 60;

Qy 350 ELHESAIVKATQISRRKHLFSRDK-----LKLFLKHQCHCEPOEGVIKIKASSLSTYKI 401

Db 484 ERHEQ-----ERRKOQLKRDQBEERRRWLKLSEERREQOER----- 521

Qy 402 AEQDPSYFFPDDPTTFIFSPANRRGRPPKRIHISQEDNVANKOTLAS---YRSKATKER 458

Db 522 REQQ-----LRREQERREQLKR-----QEEERLQORLSEQQLRQEERL 565

Qy 459 DKLKQOEMKSLAFK---AKLR---EKADALEAKKKEKEDKREKRLKIVVEERLKK 513

Db 566 EQLLKREKKEQLQERREORLKRQEERDQLLKREERROQLKREGEERL--EQRL-K 622

Qy 514 KEERLKVREKEREKUR---BEKKYVYLVKQMSKPREDMCEDDLKELPEPTPVKTR 569

Db 623 REEVERLEQE-ERRDERLKREPEERH-ELLK--SEEQERREHEQLR-EQERREQR 677

Qy 570 LPPIFGDALMWLEFLNAGELFDLQDFPDGVTLEVLSEALVGNDSGELCELLFFFLT 629

Db 678 LKRE-----EEB-----ERLEORLKREHEE-----E 698

Qy	630	AI	FQIAIBEEBEVAK	EQULTADT	TGCSLSKSLD	LSCTLSEILRLUHI	LASGADVTSANAKY	689
Db	699	RREQLAEEEOE	QARERIKSRIPK	-----	-----	-----	-----QWQLESEADARQ	740
Qy	690	RYQKRGFDAT	DDACMELRLNS	PNPSLVKLS	STSVDYLT	TPCEKMI	LHALCGKLLTLVSTR	749
Db	741	EAPQAGRAEA	POE	-----	-----	-----	-----R	783
Qy	750	DF	-----	IEDYVDILRQ	-----	AKQEFRELKAE	QHRK-----ERBEAARIRK	789
Db	784	DFTWQMAEEK	SERGRQRLSAR	PPLRQERERQLA	ERERQQRQEQ	RFLPEEEKEQ	RGQRQR	843
Qy	790	E	-----	-----	-----	EKLKEQ	BQMKKEQKLEKDEQ	829
Db	844	EREKELOF	LEBEEQ	LQRRERAQ	QOEEDGLQEDQ	ERRRQORRDQ	KWRWOL	902
Qy	830	DTSI	ESKDTEKEL	-----	-----	-----	-----	887
Db	903	RHTIYAKPAL	QEQURKQOQL	QEEBELQREER	EKRRRQBOE	-----	-----	966
Qy	888	LTAD	EEALKQEHQ	-----	-----	RKEKELLEK	QSAIACTNIPFLG-----RDMYRYWI	934
Db	961	LREER	EKRQRERQ	RYRKQK	KKLOQKEQL	-----	-----	1008
Qy	935	PPSIPGLF	IEDYSGLT	EDMLLPSPSS	FQNNVQSDPQ	VSTK-----	-----	989
Db	1009	-----	-----	-----	-----	-----	-----	1058
Qy	990	DQGRDHSV	QLPKPVH	PNRWCF	-----	-----	-----	1026
Db	1059	LQRE	RYREELQOE	EEQLGGE	RETRRRQELERQYR	KBEELQOE	EEQLREPEKR	1118
Qy	1027	NSRGHESAL	KETLLQEK	SRIC-----	-----	AOLAR	FSQE-----	1066
Db	1119	RQRE	QRCEELQOE	EEQLLREER	EKRRRQELERQYREEL	QREBELQ	RQKRYRQ	1178
Qy	1067	-----	-----	-----	-----	-----	-----	1109
Db	1179	SDLKW	PEKENAVRN	KVYCKGRE	NEQFQLED	SQV-RDRQ	SQDQLHLLGBOQER	1235
Qy	1110	QGT	LGAIKVTD	RHTWRS	-----	-----	-----	1154
Db	1236	-----	-----	-----	-----	-----	-----	1284
Qy	1155	EME	IDEOT	-----	-----	-----	-----	1212
Db	1285	EKR	ROETDRK	FREEQLQRE	EQP-----	-----	-----	1319
Qy	1213	FLKAP	LDPASD	SGRSYKTVL	DRWRESLL	SSASLSQVFL	HLSTLDRSVI	1272
Db	1320	FREEEL	LHQEGRK	FLEEEQLRE	-----	-----	-----	1359
Qy	1273	RKGDA	ENMVLCD	CGRHGHTY	CVYRPLKTV	PEGDWC	PECRPKQRCRL	1332
Db	1360	-----	-----	-----	-----	-----	-----	1380
Qy	1333	DEVED	SMGGDE	VDGDER	EGQSEEBE	EYEQEDD	DSQEEVSLPKR	1392
Db	1381	QOER	KRFR	EEQVQRQER	KERKLEB	EQQRQRHRK	KFREEBQLQRE	1440
Qy	1393	RGLSS	FSFGQOQ	QEPGRYP	SRSQSTPK	TKTVSSK	TGRSLR	1450
Db	1441	RKF	LEEBEQ	LQRQERD	-----	-----	-----	1477
Qy	1451	SRSTR	HSHG	PLQADV	FVELLS	PRKRRGR	KSANTPENS	1510
Db	1478	QQLH	QRO	-----	-----	-----	-----	1523
Qy	1511	IASK	LSQES	ESKRCR	-----	-----	-----	1560
Db	1524	EQQL	SHOERDR	KPRLB	EQKVRQ	BERKFMED	SQQLRQEGQOQL	1583
Qy	1561	VE	-----	-----	-----	-----	-----	1574

[illegible]

```
RESULT 31
US-09-949-016-7646
; Sequence 7646, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7646
; LENGTH: 2107
; TYPE: PR1
; ORGANISM: Human
US-09-949-016-7646

Query Match      3.5%; Score 305; DB 4; Length 2107;
Best Local Similarity 20.2%; Pred. No. 5.2e-13;
Matches 273; Conservative 211; Mismatches 48; Indels 380; Gaps 61;

Qy 340 FKYVQPTKELHESAIVKATQISRRHLFRDRLKFLFKHCEPQEGVVIKIKASSLSY 399
Db 125 FEYKIQ-----ELAVILKFLVHDHEDGLNLEDLENFLQ-----KAPVPST- 165
Qy 400 KIASQDFYFPDPPPTFFIPFANRRGRPPKRIHISOEDNVANKQTLASVRSKATKRD 459
Db 166 -----CSSTPPE-----LSPPHQAKREIRFLQKQVASSSSGNPNFLS-GSPSPMGD 213
Qy 460 KLLAQEEMKSLAFKAKLRKKALEAKKKEKKEKREELKIVEE----- 508
Db 214 IL-----OTQFQNRUKKQLAD-----ERSNRDELELELAENRKLUTEKDAQAMMQOR 263
Qy 509 -ERLKKKEERLKVREKEREKREKRYKVEYLKQWSPREDMECDLAKELPEPTPVK 567
Db 264 IDRLALLNEKQASPLEPKELELRDNKNESTMLHETLK-----QCODLTKESQMDRK 318
Qy 568 TRLPPEIFGD-ALMVLEFLNAGELFDLQDFPD---GVTLVLE-EALVNDSEGPLCE 622
Db 319 INQLSENGDLSFKLREFASHLQQLDALNELTBEHSKATQEWLEKQAKELKSAALQD 378
Qy 623 LFFFLFTAI FOAIAEEEBEVAKEQLTDADTKGCSI,KSLLDLSG---TLSEILRLHLASGA 680
Db 379 -----KKCLEKNEILQGLKSLQEEHLSQLQDNPPOKEGEVLGDLVQLTLEKQEA 428
Qy 681 DVTSAN-----AKYRKRGGFATDDACMELRLSNPSLVKGLSSST 721
Db 429 ATLAANNQOLQARVEMLETERGQGEAKL-LAERGHFE-----EERQOLSSLITDLS- 479
Qy 722 SVYDLTPG-----EKWKILH-----ALCGKLLTLVSTRDFTEDYVDIILROAKQFRE 768
Db 480 SISNLSQAKELEQASQAHGARLTAQVASLSELTTLNAT---IQOODQELAGLQQAQKE 536
Qy 769 LKAE--QHRKEEAAARIRKKEE---KLAEQEQKWKKEKKEKDEQBNSTADISIGE 823
Db 537 KOAQLAQTLQOQEQASQGRHQVQLSSLSKQKQQLKEVAEK-OEATRDQHAQOLATAA 595
Qy 824 EEREDFTSIESKQTEQKELQDMFTDEDEDDPGSHKGRGRRGKRGNGFKEFTREQINCV 883
Db 596 EERE---ASLRERDAALKQEA-----LEKEKAKEIL 626
Qy 884 TRELLTADEEALKQ-----EHORKEKELLEKIQSAIACTNIPFLGRDMRYRYWIFPSIP 939
Db 627 QQQLQVANEARDSAQTSVTAQREKAELSRKVEBLOAC----- 664
```

```
Qy 940 GLFTIEDYGLTDMLLPRPSSFONNVQSQDPQVSTKTGTBPLMSESTSNIDQGRDHSVQ 999
Db 665 -----VETARQEQHEAQVAELELQ-----LRSEQ----- 690
Qy 1000 LPKPVHKPNKWCIFYSSCEQLDQLIEALNSRGHRESALKETL-----LOEKSRTCAQALR 1053
Db 691 -QKATEKER-----VAQEKDQLQEQ-----QALKESLKVTKGSLKEEKKRAADAL- 735
Qy 1054 FSEKPHFSDKPDQPSFTYSRGRSSNAYDPSCWCAEQLELRDRF--LLDIEDRIYQG 1111
Db 736 --EEQRCISBELKAEATSLVEQHKRERKELEEEERAGRKGLLEARLQQLGEAHOAETEVLNR 793
Qy 1112 TLGAIKVTDRIHWSALES-----GRYELLSEENKENG-----IKTVNEDVEE-- 1155
Db 794 EL-AEAMAAQHTAESECEQLVKEVAANRERYEDSQEBAQYGFQQLMTLKECEKAR 852
Qy 1156 MEIDEOTKVIKDRLLGIKTETPTSTVNASTPOSVSVVH-YLAMALFQIEQIGIERFL 1214
Db 853 QELQE-----AKEKVAGIESHSELQISRQ-----QNELAELHANLARALQVQE-----KEV 899
Qy 1215 KAPLDASDSGRSYKTVLDRWRRESLLSSASLSQVFLHLSTL-DRSVIWSKSLINARCK-IC 1272
Db 900 RAQKLADD-----LSTLOEKMAATSKEV--ARLETLV 929
Qy 1273 RKKGDAENMVLCDGDRGHHTYCVRPKLTVPEGDWFCPCRPKQRCRRLSFRORPSLES 1332
Db 930 RKAGEQOET-----ASRELKPEPARAGDRQPEWLEEQOQGFQFCSTQ-AALQA 975
Qy 1333 DEDVEDSMGEGDEDDVGD--EEEGQSEEEYEVQEDEDDSQEEBEVS--LPKRGRPQVRL 1388
Db 976 MEREAQMGNELERLRAALMESQOQOQOER-----GQEREVARLTQERGRAQADL 1026
Qy 1389 PVK--TRGKLSSSFSSRGQOQEPGRYPSRQOQSTPTKTVSSKTCR-----SLRKINSAPP 1441
Db 1027 ALEKAARAELEMLQNALNEQ---RVFATLQEAHALTEKEGKQGLAKLGLGAAQI 1083
Qy 1442 TETKSLR-----IASRSTRSHGHFLQADVVELLSPRRKRGRKRS--ANNTPENSP 1490
Db 1084 KELEELRQTVKQLKEQLAKKEHAGS-----SGAQSEAGNTEPTGP 1126
Qy 1491 NFPNFRVIATKSEBQ-----SRSVNIASKLSLQSESKRRCRKQSPSPVTL 1539
Db 1127 KLEALRAEVSKLEQOQCKQOQADSLERSLE-ABRASRAERDSALETLQGLQLEKQAQ-EL 1184
Qy 1540 GRSSGRQGGVHLSAPEQLVVELVRHDDSW 1570
Db 1185 GHQSALASAQRELAAPRTKVQDHSKAEDW 1215

RESULT 32
US-09-949-016-7647
; Sequence 7647, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7647
; LENGTH: 2107
; TYPE: PR1
; ORGANISM: Human
US-09-949-016-7647
```

Query Match 3.5%; Score 305; DB 4; Length 2107;
Best Local Similarity 20.2%; Pred. No. 5.2e-13;
Matches 273; Conservative 211; Mismatches 487; Indels 380; Gaps 61;

QY 340 FKVKQPTKELHESAIKATQISRRKHLFSDKLLKFLKQHCPEQGVIKIKASSLSTY 399
DB 125 FEYKIQ-----ELAVILKFLVDHEDGLNLEDLENFLQ-----KAPVST- 165
QY 400 KIAEQDFYFFDDPTTFISPANRRGRPPKRIHISQEDNVANKQTILSVRSKATKRD 459
DB 166 -----CSTFPEE-----LSPSHQAKREIRFLELQKVASSSGNNFLS-GSPASPMGD 213
QY 460 KLLKQEMKSLAFERAKUKREKADALEAKKKEKEDKXKREBKIVVE----- 508
DB 214 IL-----QTPQFQMRLLKKQLAD-----ERSNRDELELAENRKLKTEKDAQIAMMOOR 263
QY 509 -ERLKKKEKRLKVEREKERELREKRYVEYLKQWSKPKREDMECDLKLPEPTPVK 567
DB 264 IDRLALLNEKQASPLEPKEBELRDNESLTMRLHETLK-----OCQDLKTEKSQMDRK 318
QY 568 TRLPPEIFGD-ALMVLEFLNAPGELFDLQDEPPD---GVTLVLE-EALVGNDSGLPCE 622
DB 319 INQLSENGDLSFKREEFASHLQLODALNELTEHSHKATQEWLEKQAKLELSAALQD 378
QY 623 LLFFFLTAIFAIAEEBEVEAKEQLTDADTKGCSLSKSLDLDSC--TLSEILRLHLIASGA 680
DB 379 -----KKCLEKNEILQKLSQLEEHLSQLQDNPQPKGEVLGVDVLQLETLKQEA 428
QY 681 DVTSAN-----AKRYQKRGCFDATTDCMELRLSNFSLVKLSST 721
DB 429 ATLAANTQLQARVEMLETEREQQEAU-LAERGHFE-----BEKQQLSLIIDLOS- 479
QY 722 SYVDLTPG-----EKKMLIH-----ALCGKLLTLVSTRDFIEDYVDILRQAKQEPRE 768
DB 480 SISNLSQAKELEQASQHGARLTAQVASTSELTILNAT---IQQQQELAGLQQAQKE 536
QY 769 LKAE--QHRKEREAAARIRKKEE--KLKEQEQMKKEQKLEKEDQRNSTADISIGE 823
DB 537 KOAQLAQTLOQEQASQGLRHQVEQLSSLSKQKEQLKVAEK-QEATRDQHAQALATAA 595
QY 824 BEREFDFTSISKOTFEKELQDQMTEDDEDDPGSHKRGGRGKRGQNGKFEFTFRQINCV 883
DB 596 EERE---ASLRERDAALKQLEA-----LEKEKAALKEIL 626
QY 884 TRELLTADDEEALKO---EHORKEKELLEKIQSAIACTNIFPLGRDRMYRYWIFPSIP 939
DB 627 QOQLQVNEARDSAOTSVTQAREKAELSRKVEELQAC----- 664
QY 940 GLFIBEDYSGLTEDMLPPSPSFQNNVQSQDPVSTKTGPELMSBSTSNIDQGRDHSVQ 999
DB 665 -----VETARQEQHEAQVAEQLQ-----LRSEQ----- 690
QY 1000 LPKPVHKNRWCIFYSCQOLDQILBALNSRGHRESALKETL-----LOEKSRICAQALR 1053
DB 691 -QKATEKER-----VAQEKDQLOQL-----QALKESLKVTKGSLSEKRRRAADAL- 735
QY 1054 FSEKFFHSDKPPQSDKPTYSRGRSSNAYDPSQCAEKOLELRDF--LLDIEDRIYQG 1111
DB 736 ---EQQRCISELKAETRLSVLEOHKREKRELEERAGRGLEARLQOLGEAHOAETEVLRR 793
QY 1112 TIGALKVTDRIHWRSALES-----GRVELLSEENKENG-----IKTVNEDVEE-- 1155
DB 794 EL-AEAMAAQHTAESCEQLKVEVAWREREDSQOEEAQYGAQFQEQMLTKEECEKAR 852
QY 1156 MEIDQTKVIVKRLGLLKTETPTSTVNASTPQSVSVVH-YLAMALFQIEQGIERRFL 1214
DB 853 QELQ-----AKEKVAGIESHSELQISRQ-----QNELAEHLANLALAQVQE-----KEV 899
QY 1215 KAPLDASDSGRSYKTVLDORWRESLSSASLSQVFLHSLTL-DRSVMSKSLINARCK-IC 1272
DB 900 RAQKLADD-----LSTLQEKMAATSKEV--ARLETLV 929

1273 RKKGDAENMVLCDGCDRGHHTYCVRPKLVKTVPEGDMFCPECRPKQRCRRRLSFRQPSLES 1332
DB 930 KRAGEQOET-----ASRELKVEPARAGDRQPEWLEEQGRQFCSTQ-AALQA 975
QY 1333 DEDVEDSGGDEVDGD--EEGQSEEEVEVEQDEDDSOEEERVS--LPKGRPOVRL 1388
DB 976 MERAEQNGNELRUALMESQGOQOQEEER-----GOQEREVARLTQERGAQADL 1026
QY 1389 PVK--TRGKLSFSSFRSQOQPGRYPSRSQOQSTPKTTVSSKTGR-----SLRKINSAPP 1441
DB 1027 ALEXAARAELEMLRQNALNEQ---RVEFATLQEAALAHALTEKQKQDELAKLGLAAQI 1083
QY 1442 TETKSLR-----IASRSTRSHGPLQADVVELLSPRKRGRKIS--ANNTPENSP 1490
DB 1084 KELEELROTQVKLEQLAKKEKHAAG-----SGAQSEAAGRTEPTGP 1126
QY 1491 NFPNFRVIATKSSSQ-----SRSVNIASKLSQSESKRCKRCKRQSPSPVTL 1539
DB 1127 KLEALRAEVSKLEQCCQKQEQADSLERSLE-AERASRAERDLSALETQGLLEKQAQ-EL 1184
QY 1540 GRRSSGRQGVHLSAFEQLVVELVRHDDSW 1570
DB 1185 GHSQALASAQRELAARFTKVQDHSKAEDW 1215

RESULT 33
US-08-466-390-4
; Sequence 4, Application US/08466390
; Patent No. 5686562
; GENERAL INFORMATION:
; APPLICANT: TOURKATLY, GARY
; APPLICANT: LIDGARD, GRAHAM P
; TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
; TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TESTA, HURWITZ & THIBEAULT
; STREET: 125 HIGH STREET
; CITY: BOSTON
; STATE: MA
; COUNTRY: USA
; ZIP: 02110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/466,390
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: PITCHER ESQ, EDMUND R
; REGISTRATION NUMBER: 27,829
; REFERENCE/DOCKET NUMBER: MTP-013
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 248-7000
; TELEFAX: (617) 248-7100
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2101 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-466-390-4

Query Match 3.4%; Score 297; DB 1; Length 2101;
Best Local Similarity 20.0%; Pred. No. 2e-12;
Matches 271; Conservative 210; Mismatches 486; Indels 388; Gaps 61;

QY 340 FKVKQPTKELHESAIKATQISRRKHLFSDKLLKFLKQHCPEQGVIKIKASSLSTY 399
DB 119 FEYKIQ-----ELAVILKFLVDHEDGLNLEDLENFLQ-----KAPVST- 159


```
Qy 623 LLFFFLTAIFAIAEIEEVEEVAKEQOLTADTKGCSLSKSLDLDSC--TLSEILRLHLASGA 680
Db 373 -----KKCLEKNEILQGLKSLQEEHLSQLODNPPEKEGVLGDVQLQLETKQEA 422
Qy 681 DVTSAN-----AKRYQKRGGFATDDACMELRLSNPSLVKKLSST 721
Db 423 ATLAANNQOLQARVEMLETEREQOEAKL-LAERGHFE-----EEKQLSSLTDLQS- 473
Qy 722 SVYDLTPG-----EKMKILH-----ALCGKLLTLVSTDRFDIEDYVDIILROAKQEPRE 768
Db 474 SISNLSQAKELEQASQAHGARLTAQVASLSELTTLNAT---IQQQDOELAGLQKQAKE 530
Qy 769 LKAE--QHRKEREREAARIRKKEE---KLKEQKMKKEKOEKLKEDQORNSTADISIGE 823
Db 531 KOAQAQLQOQEQSQSLRQVQLSSSLKQKQOLKEVAEK-QEATRDQHAQOLATAA 589
Qy 824 EEREDFTSIESKOTEQEKELQDMFTDEDEDDPGSHKRRGRKGQNGKFKEFTROEQINCV 883
Db 590 EERE---ASLRERDAALKQLEA-----LEKEKAUKLEIL 620
Qy 884 TRELLTADEEALKQ-----EHORKEKELLEKIQSAIACTNIPPLGRDRMYRRYIPFSP 939
Db 621 QQQLQVANEARDSAQTSVTQAREKAELSKEVELOAC----- 658
Qy 940 GLFTEEDYGLTDMLLPRPSSFQNNVQSDPOVSTKTGTEPLMGSESTSNIDQGRDHSVQ 999
Db 659 -----VTAQOEQHEAQVAEELQ---LRSEQ----- 684
Qy 1000 LPKVPKPNRWCFYSSCQDLQLEALNSRGHRESALKETL-----LOEKSRIACAQLAR 1053
Db 685 -QKATEKER-----VAQEKDQLEQ-----QALKESLKVTKGSLEEKRAADAL- 729
Qy 1054 FSEKHFHSDXPQDPDKTYGRSSNAYDPSQCAEKQLELRDP--LLDIEDRIYQG 1111
Db 730 --EQQRCISLKAETRLVQHKRERKELEERAGRGLEARLLQLGEAQAETEVLRR 787
Qy 1112 TLGAIK-----VTDRIHW-----SALESRYELLSEENKENGIIKTVEDV 1153
Db 788 ELAEMAAQHTAESECEQLKVEAAMRGYEDSQOEAEQYGMFQEQ-----LMTLKEEC 842
Qy 1154 EE--MEIDEQTKVIVKDLGKITETPSTVNASTPQSVSSVHV- YLAMALFQIEQIE 1210
Db 843 EKARQELQEQ-----AKEKVAGIESHSELQISRQ-----QNKLAELHANLARALQVQE--- 890
Qy 1211 RRLFKAPLADSDSGRSYKTVLDRWRESLLSASISQVFLHLSLTL-DRSVIWSKILNARC 1269
Db 891 -KEVRAQKLADD-----LSTLQEKMAATSKEV--ARL 919
Qy 1270 K-ICRKGDAENMVLCDGCDRGHHTYCVRPKLKTVPEGDWFCPCRCRQKORCRLSFRQRP 1328
Db 920 ETLVRKAGEQET-----ASRELKPEPARAGDQRPWLBSEQQGRQPCSTQ-A 965
Qy 1329 SLESDVEDSMGGDEDDVDGD--EEGQSEEEYEVEQDEDDSQEEBEVS--LPKGRGP 1384
Db 966 ALQAMERAEQMGNELEFLRAALMESQOQOEER-----QOQEREVARLTOERGRA 1016
Qy 1385 QVRLPVK--TGKLSSSFSRGQOQEPGRYPSPRSQQSTPTKTVSSKTR-----SLRKIN 1437
Db 1017 QADLALEKAARAELEMRQLQNALNEQ---RVBFATLQELALAHALTEKEGQOQELAKRLGLE 1073
Qy 1438 SAPPTETKSLR-----TASRSTRHSHQPLQADVVELLSPRRKRGRKS--ANNTP 1486
Db 1074 AAQIKELBELQTVKQLKEQAKKEKHAHG-----SGAQSSAARGTE 1116
Qy 1487 ENSNFPNFRVIATKSSSQ-----SRSVNIASKLSQESKRRCKRQSPSPS 1535
Db 1117 PTGPKLEALRAEVSKLEQCCQKQEQADSLRSLE-AERASRAERDSALETLQOQLEKA 1175
Qy 1536 PVTLGRSSRGQGGVHLSAFEQLVVELVRHDDSW 1570
Db 1176 Q-ELGHSQSALASAQRELAAPRTKVDHSKAEDEW 1209
```

```
RESULT 36
US-08-195-487-4
; Sequence 4, Application US/08195487
; Patent No. 5783403
; GENERAL INFORMATION:
; APPLICANT: TOURKATLY, GARY
; APPLICANT: LIDGARD, GRAHAM P
; TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
; TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TESTA HURWITZ & THIBEAULT
; STREET: 53 STATE STREET
; CITY: BOSTON
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/195,487
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/901,701
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: PITCHER ESQ. EDMUND R
; REGISTRATION NUMBER: 27,829
; REFERENCE/DOCKET NUMBER: MTP-013
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/248-7000
; TELEFAX: 617/248-7100
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2101 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-195-487-4
```

Query Match 3.4%; Score 297; DB 1; Length 2101;
Best Local Similarity 20.0%; Pred. No. 2e-12;
Matches 271; Conservative 210; Mismatches 486; Indels 388; Gaps 61;

```
Qy 340 FKYKVQPTKXELHESATVKTQISRKHLSRDKLKLFLKQHCPQEGVIKIKASSLSY 399
Db 119 FEYKIQA-----ELAVILKFDHEDGLNLEENFLQ-----KAPVPSI- 159
Qy 400 KIAEQDSYFPDPDPTTFIFSPANRRGRPPKRIHISQEDNVANKOTLASYSKATKRD 459
Db 160 -----CSSIFPEE-----LSPPSHOAKREIRFLEQKVASSSSGNNFLS-GSPASPMGD 207
Qy 460 KLLQKQEMKSLAFKAKLKREKADALEAKKKEKEDKCKRRELKIVVEE----- 508
Db 208 IL-----QTPQFQMRLLKQLAD-----ERSNRDELLELAENRKLKLTEDKDAIAMMOOR 257
Qy 509 -ERLKKKEERLKVEREKEREKREKRYVEVLKQWSKPRDEMCDDDLKEPPEPVK 567
Db 258 IDRLALLNEKQASPLKPEKLEELDRKNESITWKLHETLK-----QQDCLKTEKSQMDRK 312
Qy 568 TRLPPEIFGQ-ALMVLFEFLNAFGELFDLQDEFPD---GVTLFVLE-EALVGNDSGEPICE 622
Db 313 INQLSENGDLSFKLREFASHLQQLQDALNELTTEHKSATQEWLEKQAKLEKLSAALQD 372
Qy 623 LLFFFLTAIFAIAEIEEVEEVAKEQOLTADTKGCSLSKSLDLDSC--TLSEILRLHLASGA 680
Db 373 -----KKCLEKNEILQGLKSLQEEHLSQLODNPPEKEGVLGDVQLQLETKQEA 422
```


Qy	1270	K-ICRKKGDANMVLCDGCDRGHHYTCVRPKLKTVPEDGMFCPECRPKQRCRLSPRQP	1328
Db	920	ETLVRKAGEQOET-----ASRELVRKEPARAGDQPEWLEEQGRCQFCSTQ-A	965
Qy	1329	SLESDDEDVSMGGEDEVDGD--EEGQSEEEYEVEQEDDSQSEEEVS--LPKRG	1384
Db	966	ALQAMEREAEQMGNELRLAALMESQGOOEER-----GQEREVARLUTQERGA	1016
Qy	1385	QVRLPVK--TRGKLSSFSRGGQOQFGRVPSRQOSTPKTTVSSYKGR-----SLRKIN	1437
Db	1017	QADLALKAARALEMLQNALNEQ---RVFATLQALAHATEKGQOELAKRLGLE	1073
Qy	1438	SAPPTETKSLR-----IASRSTRHSGPLQADVVELLSPRRKRRGRKS--ANNTP	1486
Db	1074	AAQIKELLELRQTVKQLKEQLAKEKEHAG-----SQAQSEAAQRT	1116
Qy	1487	ENSPNFFNFRVIATKSSEQ-----SRSVNIASKLSQSESKRRCKRQSP	1535
Db	1117	PTGPKLEALRAEVSKLQOQCKQOQADSLERSLE-AERASRAERDALETLQOOLEKA	1175
Qy	1536	PVTLGRSSRGQGVHELSAFQVLVRLVRHDSW	1570
Db	1176	Q-EIGHQSALASAQRELAFAFTKVODHSKAEDEW	1209

RESULT 40

```

RES001.40
US-09-595-684B-31
; Sequence 31, Application US/09595684B
; Patent NO. 6544766
; GENERAL INFORMATION:
; APPLICANT: Beraud, Christophe
; APPLICANT: Ohashi, Cara
; APPLICANT: Sakowicz, Roman
; APPLICANT: Vaisberg, Eugen
; APPLICANT: Wood, Kenneth
; APPLICANT: Yu, Ming
; TITLE OF INVENTION: Human kinesins and methods of producing
; TITLE OF INVENTION: and purifying human kinesins
; FILE REFERENCE: cytop036
; CURRENT APPLICATION NUMBER: US/09/595,684B
; CURRENT FILING DATE: 2002-06-24
; PRIOR APPLICATION NUMBER: 09/295,612
; PRIOR FILING DATE: 2000-04-20
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 31
; LENGTH: 2662
; TYPE: PRT
; ORGANISM: Human
US-09-595-684B-31

```

Query Match	3.4%	Score 293.5	DB 4	Length 2662
Best Local Similarity	18.7%	Pred. No. 5.2e-12		
Matches 305	Conservative 287	Mismatches 592	Indels 449	Gaps 74
Qy	212	LTYQEALESEKARQNLSQPEPELIIPVLYTSL-THR-----SRLEHCDDIPAVVXDY	266	
Db	1143	LNQVEENSEQKINEIENLKNELTLEHMETERLELAQKLNENYEVKSITK	1202	
Qy	267	FVEETVEIRNNGARLOCTILEV-----LPPSHQNGFANGHV-----NSVDGETI	311	
Db	1203	VKLQELQSFETERDHLRGYIREIBATGLQTKELKIAHIHLKEHQETIDELRRSVSEKTA	1262	
Qy	312	IISDDSDSETQSCSFQ-----GKKDAIDPLLPKYVQVPTKBL-----HESAIKVATQI	362	
Db	1263	QIINTQDLKSHTKLQEEIPVLHBEQELLPNVKVSVQETMNELELLTQSTTKDSTTL	1322	
Qy	363	SR-----RKHLFSRDKLKLFLQ--HCEPOEGVIKIKASSLSY-----KIAQD	405	
Db	1323	ARTEMERLRNLNEKFEQSEIKSUTKRDNLTKTKEALVGHQDLKEHIRETLAKIQES-	1381	
Qy	406	FSYFFPDPTTFISPNRRRRGPPKRIHISQEDNVANKOTLASYSRKATKDERKLLQOE	465	

D	b	1382	--	---	----	-----	-QSKOEQSLNNKKDKDNETTK--	I-VS-EMEQFKPKDPSALLRI	1418
Q	y	466	EMKSLAF	EKA	KREKADALEAKKEKEDKEKRE-	----	-----ELUKTIVBERLKKKEEKE	518	
D	b	1419	EIENMGL-	-SKRLQESHDEMKVASKDDLRLOEVLSQESDQLKENIKIVAKHLETEE	1476				
Q	y	519	RLKVYR-	-----EKERE-	-----	-----KLRE-EKRKYVEY	541		
D	b	1477	ELKVAHC	LKEQEQTINELRVNLSEKETETSTIQKQLEAINDKLQNKIQBIYEKEEQLN-	1535				
Q	y	542	LKQWKS	PREDMECDLKPPEPVPVKTRLPPFIFGDALMWLE-F	FLNAPGELPDODE--	597			
D	b	1536	IKOIS	VOENV-NELKQFKEHRKAK-----DSALQSI	ESKMSELTVRLQESQEIQ	1585			
Q	y	598	--FPDGVT	LEVLEAL-----VGNDSEGPLCELLFFFTAITAIFAQIAABE-EE	640				
D	b	1586	IMIKEEM	KRVQEAQIERDOLKENTKEIVAKMESQEQYOPLKMTAVNETQKWCEI	1645				
Q	y	641	EVAKEQ	TADATKGCSLKLDLSCTUSEILRHILASGADVTSANAKRYQKRGGFDT	700				
D	b	1646	EHLKEQP-	-ETOKLANLETETENIRLTQTLIHEN-LBEMRSVT-----KERDDLRVS	1693				
Q	y	701	DDACWE	LRLSNPVLKLSVSVVLTTPGKWKILH-----ALCGKLLTLVSTR----	749				
D	b	1694	EET---	LKVRDQKWLRTITTRDLQEBELKIVMHLEKHQBETIDKLGI	VSEKTNEI	1750			
Q	y	750	----DFIED	VDIILR-----QAKOFRELKAEHQHREEAAAIR-----KRKEKCLKBOEQ	798				
D	b	1751	SNMQOLE	HNSDALKAQDLKIQEELR-IAHMLKEQOETIDKLRGIVSEKTDKLSNQK	1808				
Q	y	799	-----XMK	EKOBLKEDSORNSTADISGEER-----EDFPTSIESKDTE	839				
D	b	1809	DLENSA	NKLOEQIELKANEHQITLTKKQVNETQKQVSEMEOQLKQIKDOSSLTSLKLEIE	1868				
Q	y	840	QKELDQ	DMFTEDEDPCSHKRRGRKRGONGPK-----EFTROEIOINC	VRTLELT	889			
D	b	1869	NLNLAQ	BELHENLEMBSVMKERNLRAVESTLKLERDQLESQETKARDLE-IQOBLKT	1927				
Q	y	890	ADEEAL	QBOHORKEBELBKISOAIACTNI PPLGRDRMYRYYWI PPSIPCLFI	EEDYSYG	949			
D	b	1928	A---RMLS	KEH---KETVDKLERIKSEKTI-----QISD	1955				
Q	y	950	LTEMPL	PRSSFOQNVQSODPOQVSTKTGPMLSESTSDNOQPRDHSVOLPKPVHPNR	1009				
D	b	1956	IQDOL-	-----DKSKDELQKIQLOKQELQOLLRVKVEDVNNSHKKNIMEOULKQK-P	FEPNY	2009			
Q	y	1010	WCFYSC	-----OLDOLIEALNSRGHRESAL---KETLLQEKERI	CAQLAR-PSEE	1057			
D	b	2010	LC---K	CMDNFQTLKXLESLEIRIVAKERDELRI KESLKWEROQFATLAREMIARD	2066				
Q	y	1058	KTFHSP	KDPQPSKPTYSRGRSSNAYDPSQCAE-KQLELR-----LRDPLLDIED	1106				
D	b	2067	RQHNVKPE	---KLLSDGOQHLMESIREKCSRICKELLKEYSEMDDHYECLNRLSLDLEK	2123				
Q	y	1107	RI-YQG	TILGAIKYTDRHIWESALES---GRYEL--LSEENKENG I---IKTVNEDEV-	1154				
D	b	2124	EIEPHR	IMKKLVVLSVTVIKEQHECINKPFMDFIDEVEKOKELLIKIQHOQQDCDVP	2183				
Q	y	1155	-----E	MEIDEOTKVIIVKDRLLAGITETPTSTVNTASTPQSVSSVVHYLAMA	1201				
D	b	2184	SRELRL	KLQNOMDLHIEELIKOPSESEFPFSIKTEFOQVLNKRKEMTQFLEEMLN---	2239				
Q	y	1202	LFQIEQ-	---GIER-----RFLKAPLDDASDGRSYKTVLDRWRESILLSSASL	1244				
D	b	2240	RFDIEK	LNGIKQENDRICQVNNFNFRRIITAINNESTEFEERSATISKEWEOQLKSUKEK	2299				
Q	y	1245	SO-VFL	HLSTLDRSVIWSKSIILNARCKICRKKGDAENMVLCDGDCGRGHHTYCVRPKLKT V	1303				
D	b	2300	NEKL	FKNYQILTYSL-----ASGAVNPPTQD--NKNPHTVTSRATQILT--	2341				
Q	y	1304	PEGDW	FCPECRPKQRCRRLS-----FRORPSILEDEDDVEDSNSGGBEDE	1346				


```

Db 2342 -----EKIRELENSLHEAKSAMHKSIIKMQKELEVTNDIIAKLOAKVHE 2388
Qy 1347 VGDSEEGOSBEEVEYQDE-----DQSEEEVSLPKRGQVRLPVKTRGKLSS 1398
Db 2389 SNKCLEK--TKETIQVLQDKVALGAKPYKEBIEIDKMKLVKIDLEKMKNAKEPPEKEISA 2445
Qy 1399 SPSSRGQOQEPGRY-----PSRSQOSTPTKTVSSKTRSLRKINSAPPT-----ETK 1445
Db 2446 TKATVEYQKEVIRLRENLRRSQQAQDTSVISEHTD---POPSNKLTCGGSGIGVQNTK 2502
Qy 1446 SRIASRSTRSHGLOADVFE-----ELLSPRR-----KRRG 1478
Db 2503 ALLIKSEHR-----LEKEISKLQKQNEQLIKQKNEILLSNNOHLSNEVKTWKERTLKREA 2557
Qy 1479 RKSANNTPNSPNFNRVIATKSEQSRSRVNIASKLSQE-----SESKRRC-----RKQOS 1531
Db 2558 HKQV--TCENSPKSP--KVTGTASKKKQITPSQCKERNLQDPVPKESPKSCFFDSRSKSL 2613
Qy 1532 PEPSPVTILGRSS 1544
Db 2614 PSPHPVRYFDNSS 2626

RESULT 41
US-09-538-092-1252
; Sequence 1252, Application US/09538092
; Patent No. 6753114
; GENERAL INFORMATION:
; APPLICANT: Giot, Loic
; APPLICANT: Manefield, Traci A.
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538,092
; CURRENT FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CuratSeqFormatter Version 0.9
; SEQ ID NO 1252
; LENGTH: 2663
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (0)-(0)
; OTHER INFORMATION: Polypeptide Accession Number Q02224
US-09-538-092-1252

Query Match 3.4%; Score 293.5; DB 4; Length 2663;
Best Local Similarity 18.7%; Pred. No. 5.2e-12;
Matches 305; Conservative 287; Mismatches 592; Indels 449; Gaps 74;

Qy 212 LTYQALSEKAKARNLQSFPELPIPVLYTSL--THR-----SRLHEICDDIFAVKDRY 266
Db 1144 LNVQSEMOMKINEIENLKNELKNKELTLEHMETERLELAQKLNENYEVKSTIKRK 1203
Qy 267 FVEETVEIRNNGARLQCTILEV---LPPSHQNGFANGHV-----NSVDGETI 311
Db 1204 VLKELQKSFETERDHLRGVIREATGLQTKBELIAHILKHEQHTIDELRRSVSEKTA 1263
Qy 312 IISDSDSETQSCSPQN-----GKKDAIDPLLPKYKVQPTKKEI-----HESAIKATQI 362
Db 1264 QIINTQDLKSKTKLQEBEIPVLHBEQELLPNVKVYSEVQETMNELELLTQSTTKDSTTL 1323
Qy 363 SR-----RKHLFSDRKDKLFQK--HCEPQBGVIKIKASSLSTY-----KIAEOD 405
Db 1324 ARIEMERLRLNEKFQESBEIKSLTKERDNLTKIKEALEVKHDQLKEHIRETLAKIQES- 1382
Qy 406 FSYFPDPPPTFIIPANRRGRPPKRIHISQEDNVANKQILASYSRATKVERDKLLKQE 465
Db 1383 -----QSKQSQSLNMKKEKDNETTK--IVS-EMEQFKPKQSALLRI 1419

```

```

Qy 466 EMKSLAFKAKLKREKADALEAKKKEKEDKEKKRE-----ELKKIVVEERLKKKESKE 518
Db 1420 EIEMLGL--SKRLOESHDEMKSAKEDDLQRLQEVLOSQDLKENIKKIVAKHLETEE 1477
Qy 519 RLKVER-----EKERE-----KURE--EKKKYVEY 541
Db 1478 ELKVAHCLCKQEETINELRVNLSEKETEITIQKLEAINDKLNKIQEYIEKEEQNL- 1536
Qy 542 LKQMSKPREDMECDLPEPTPVKTRLPPEIPFDALMWLE--FLNAFGELFDLQDE-- 597
Db 1537 IKQISEVQENV--NELQPKHEKAK-----DSALQSIESKMLELTNRLQESQEEIQ 1586
Qy 598 --FPDGVTVLEBAL-----VGNDSGELPCCELLPFLPTAIQAIABE--BE 640
Db 1587 IMIKEEMKRVQALQIERDQLKENTKEIVAKMKESQEKYQFLKMTAVNETQKMCBI 1646
Qy 641 EVAKEQLTDATKCSLKSLLDSCSLSEILRLHLASGADVTSANAKRYVKRGGFAT 700
Db 1647 EHLKEQF---ETQKLNENIETENIRLTQILHEN--LEEMRSVT-----KERDILRSV 1694
Qy 701 DDACWELRLSNPSLVKKLSSTSVYDLTPGEMKILH-----ALCGKLLTLVSTR--- 749
Db 1695 EET---LKVERDQKLENLRETIITRDLEKQEBELKIVHMLKHEQETIDKLRGIVSEKTEI 1751
Qy 750 ---DFIEDYVDILR---QAKQBFRELKASQHKRKEEAAAARIR--KRKEEKLUKEQEQ 798
Db 1752 SNMQKDLHSNDALKAOCLKIQEBELR--IAHMLKEQEQETIDKLRGIVSEKTDKLSNMOK 1809
Qy 799 -----KMKKEQKLEDEQORNSTADISICEEER-----EDPDTSIESKOTE 839
Db 1810 DLENSNAKLOEQIELKANEHQILITLKQVNETOKKVSSEMOLKQIKQISLUSKLBIE 1869
Qy 840 QKELQDMFTEDDDPGSHKGRGRKGQNGFK-----EFTQEQNCVTRRELT 889
Db 1870 NLNLAQELHENLEEMKSVKERNLRRVEETLKLERDOLKESLQETKARDLE-IQELKT 1928
Qy 890 ADEEALKQEHORKEKELLEKIQSAIACNTNIPPLGRDRMYRRYWFPSIPGLFIEEDYSG 949
Db 1929 A---RMLSKEH---KETVDKLRKISEKTI-----QISD 1956
Qy 950 LTEDMLLPSPSSFQNNVQSDPQVSTKTGTEPLMSESTSNIDQGPDRDHSVQLPKPVHKNR 1009
Db 1957 IQKDL-----DKSKDELQKIQELQKELQLLRVKNEDVNMSHKKINEMEWLKKQ-FBPNY 2010
Qy 1010 WCFYSSCE-----QLDQLEALNSRGHRESAL---KETLLOEKSRICAQALAR-FSEE 1057
Db 2011 LC---KCEMDNFQITKKLHESLEIRIVAKERDELRRIKESLKWERDQFIATLREMIARD 2067
Qy 1058 KFHFSKPDQPSKPTYSGRSSNAYDPSQMACB-KQLELR-----LRDFLLDIED 1106
Db 2068 RQNHQVKE---KRLSDGQQLHESLREKCSRIKELLKRYSEMDDHYECLNRLSLDLEK 2124
Qy 1107 RI-YQGTILGAIKVTDRHIWRSALSS---GRYEL--LSEENKENG---IKTVNEDVE-- 1154
Db 2125 EIEFHRIMKKLYVLYSVYTKIEEHCINCINCFDIDEVEKQKELLKIKIHLQOODCQVP 2184
Qy 1155 -----BWEIDEQTKVIVKORLLGKITETPTSTVSTNASTPOSVSVVHYLAMA 1201
Db 2185 SRELRDLKLNQMDLHIEILKDFSEFPPIKTEFQVVLNRRKEMTOFLSEWLN---T 2240
Qy 1202 LFOIEQ---GTER-----RFLKAPLDASDSGRSYKTVLDRWRRESLSSASL 1244
Db 2241 RFDIEKLNKGIQKENDRICQVNNFNRRIIAII MNESTEFERSATISKWEQDLKSLKEK 2300
Qy 1245 SQ-VFLHLSTLDRSVINWKSILNARCKICRKKGAENNVLCGDCGRGHTTVCVRPKLTV 1303
Db 2301 NEKLFQNYQTILKSL-----ASGAQVNPFTQD--NKNPHVTSTRATQLT-- 2342
Qy 1304 PEGDWFCPCRPKQRCRLS-----FRQRPSSLESDESDVEDSMGGEDE 1346
Db 2343 -----SKIRELENSLHEAKSAMHKSIIKMQKELEVTNDIIAKLOAKVHE 2389

```

Qy	1347	VDGDEEEQSSSEEEYEVEQDE-----DSDQEEVEVSLPKRGSPQVRLPVKTRGKLSS	1398
Db		: : : : : : : : : : : : : : : : : : : :	
Db	2390	SNKCLEK---TKETIQVLQDKVALCAKPYKEBIEIDLKMKLVKIDLEKMKNAKEFEKEISA	2446
Qy	1399	SFSRGGQOEGRY-----PSRSQQSTPTTYSSTGKRSRLRKINSAPPT-----ETK	1445
Db		: : : : : : : : : : : : : : : : : : : :	
Db	2447	TKATVEYQKEVIRLLRENLRESQAQDTSVISEHTD---PQPSNKP LTCGGSGGIVQNTK	2503
Qy	1446	SLRIASRSTRSHGFLQADVVF-----ELLSPRR-----KRRG	1478
Db		: : : : : : : : : : : : : : : : : : : :	
Db	2504	ALILKSEHR-----LEKEIKLKQONQLIKQKNELLSNNQHLNSNEVKTWKERTLKREA	2558
Qy	1479	RKSANNTPENSPNPNFRVIATKSEQSRSVNIASKLSQE---SESKRRC-----RKQOS	1531
Db		: : : : : : : : : : : : : : : : : : : :	
Db	2559	HKQV--TCENSPKSP--KVTGTASKKKQITPSQCKERNLQDPVPKESPKSCFFDSRSKSL	2614
Qy	1532	PEPSPVTLGRSSS	1544
Db		: : : : : : : : : : : : : : : : : : : :	
Db	2615	PSHPVRYFDNSS	2627
RESULT 42			
US-09-976-594-726			
; Sequence 726, Application US/09976594			
; Patent No. 6673549			
; GENERAL INFORMATION:			
; APPLICANT: Furness, Michael			
; APPLICANT: Buchbinder, Jenny			
; FILE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS			
; TITLE REFERENCE: PA-0041 US			
; CURRENT APPLICATION NUMBER: US/09/976,594			
; CURRENT FILING DATE: 2001-10-12			
; PRIOR APPLICATION NUMBER: 60/240,409			
; PRIOR FILING DATE: 2000-10-12			
; NUMBER OF SEQ ID NOS: 1143			
; SOFTWARE: PERL Program			
; SEQ ID NO 726			
; LENGTH: 2468			
; TYPE: PRT			
; ORGANISM: Homo sapiens			
; FEATURE:			
; NAME/KEY: misc feature			
; OTHER INFORMATION: Incyte ID No. 6673549 4295277CD1			
US-09-976-594-726			
Query Match 3.3%; Score 288.5; DB 4; Length 2468;			
Best Local Similarity 19.4%; Pred. NO. 1.1e-11;			
Matches 304; Conservative 219; Mismatches 547; Indels 497; Gaps 67;			
Qy	210	PGLTVQEALESEKKARONLQSF--PEPL-----IIPVLVLTSLTHRSRLHEICDDIF	259
Db		: : : : : : : : : : : : : : : : : : : :	
Db	370	PNIKKKISIEACTFLQYLNLKLSMKPELFRSVGNITDPVILFQKMGVGLKEMVYLPVK	429
Qy	260	AYVKDRYFVE-----ETVEVIRNNGARLOCTILEVLPPSHQNGFANGHVNSV-----	306
Db		: : : : : : : : : : : : : : : : : : : :	
Db	430	SSKEWQYFMQWGTGNKDKAFILPNGQVD-----LPISYLTLS-----VSSLIVHWA	478
Qy	307	-DGETIISDDSETQSCSFQNGKKDAIDPLLFYKVKQP--TKKELH-----ESAIVKAT	360
Db		: : : : : : : : : : : : : : : : : : : :	
Db	479	NPAEKIIRVLFPGNSTQYNILEGLEKLHLD-----FLKQPLATQKDLTGQVTPVVRQT	533
Qy	361	QISRKKHLFSDKL-----KLFLQHCPEQBGVLKIKASSILSTYKIAEQDPSFYFP	411
Db		: : : : : : : : : : : : : : : : : : : :	
Db	534	KLKQAD--SRESLKPAAKPLPSVSRVSKESKEETPEVTKVNHVE--KPPKVSSEKQVMVK	590
Qy	412	DDPTTFIFSPANRRRGPPKRIHISQEDNVANKQTLASYRSKATKE-----RDKLLKQEM	467
Db		: : : : : : : : : : : : : : : : : : : :	
Db	591	DKPVKTEKPSVTEKEVPSKEEPSVKAIEAKQA--TDVKPKAAKEKTVKKEKVKPKEDK	649
Qy	468	KSLAFEAKLKRE---KADALEAKKKKEDKKKREELKKIIVEERLKKKBEKERLKYE--	523
Db		: : : : : : : : : : : : : : : : : : : :	
Db	650	KE---EKEKPKKVAKKEDKTPIKKEKPKKEEVKVKGIKKE--EKKGPKEKVKKET	704

QY	524	--REKEREKUREKRYEVLKOWSKPRDEMCDDLKELPE-----PTPVKTLRPPPEIFG	576
DB	705	PKPEVKKEVKEEKK---EVKKEKEPKK---EIKKLPKDAKKSSTPLSEAKKP-----	752
QY	577	DALMVLFLNALFGLFDLQDFPDGVTLEVLBEALVGNDESEGLCELLFFLFTAIQAIA	636
DB	753	-----AALPKPVP	760
QY	637	EEBEVAKEOLTADTKGCSLKSLDLDSCSLRILHILASGADVTSANAKVRYOKRGG	696
DB	761	KKEESVKKDSVAAGKPK-----EKGIKVIKKEGKAABA-----AANG	800
QY	697	FDATDDACMELRSLNPSLVKKJLSTSYVOLTPEKMKILHALCGKLLTLVSTRDFIE---	753
DB	801	TGATTAAMV-----AAAGIAIGPAKELEAERSLWSSPEDL--TKDFEELKA	845
QY	754	DYVDILQAQOEFRELKABOHRKEREBAAARIKKEEKUK-----EEOQKM	800
DB	846	BEVDVTXDIKPQLELIBDESKJLETPEVAVVIQKEREVTKGAESPDEGITTTEGEC	905
QY	801	KEKEKELKEDEORNSTADISIGEEERDDPOTSXKDEKELDQDMFTDEDEDPCSHKR	860
DB	906	EQTPEEL-EPVEKGVDDIIEKFEDEGAGFESSETGYEEK-AETEEABEPEDGEHEVC	963
QY	861	GRRKRGQNGFKEFTREQINCY--TRELLITA-----DEEAL-KOEHRKEKELLEK	910
DB	964	VSASKHSPTDEESAKAEADAYIREKESVASGDDRAEDMDRAIEKGEAEQSEERADEE	1023
QY	911	IQSAIACTNIPPLGRDRMYRYWIFPSIPGLFTIEDY-----SGLUTDM--LLPR	958
DB	1024	DKABDA-----REBEYE-----PEKMEADYVMAVDKAAEAGAEQYGFLLT	1067
QY	959	PSSFQNNVQSDPQVSTKTCEPLM-----SESTSNIDQGRDHSVLQPKPVHKPNWCIFY	1013
DB	1068	PTK-OLGAQSPGREPASSIHDETLPGGSESEATASDEENR-----QPEFTAT	1116
QY	1014	SSCEQ-----LDQLEALNARGHRESALKETLQEKSRICQAQLARFSEEFHFS	1062
DB	1117	SGYTQSTIEISSEPTPMENSTP-----RDVMSDETNNEETESPQEFVNIIT--KVES	1168
QY	1063	DKPOPDSKPT-----YSRGRSNAYD-----PQMCAGKOLELRLRD---	1099
DB	1169	LYSQEYKPADVTPLNGFSBGSKTDATDGRDYNASASTISPPSMBEDKFSRGAARDAYC	1228
QY	1100	-----FLLDIENRIYOGTLGAIKVTDRH-----IWSAL--ESGRYELLSEEN	1140
DB	1229	SEVKASTTLDIKOSI--SAVSSEKVSPSKSPSUSPPSPLEKTLGERSVNFSLTPNEI	1286
QY	1141	KENGIIKTVNEDEVEMEID-EQTKVIVKDRLLGKTETPTSTVSTNASTPOSVSVVHYLA	1199
DB	1287	K-----VSAEAEVAPVSEVTQEVVEHCASPEDKTLFVSPQSVCAGHTPY--	1337
QY	1200	MALFQIEQGIERRFLKAPLDASDGRSXYKTVLDRWRESLLSSASLSQVFLHLSTLDRSVI	1259
DB	1338	-----QSPTDEKSS-----HLPT--EVI	1353
QY	1260	WSKSIILNARCKICRKKGDAAENWVLDCDCDRGHHTYCYVRPKLXTVPBGDMFCPECRPKQRC	1319
DB	1354	EKPPAVPVSPFSDAKDENERA-----SVSPMDPEVPDSE-----SPIEKV	1394
QY	1320	RRLSFRORPSLESDVEDSMGGEDDEVDG-----	1349
DB	1395	--LSPLRSPLIGESAYESFLNDDKASGRGAESPKEESKGQSPDQVSPVSEMTSTS	1452
QY	1350	---DEEBQSEEE-----EYEEVEQDEDDQOEBEVSIPKRGRPOVRLPVKTRGKLSSFS	1401
DB	1453	LYODKQSGKSTDPAPIKEDFQGEKKTDDVE-----AMSSQPALALDERKGLDVSPQOI	1505
QY	1402	SRGOQEPGRYPSPRSQOSTPKTTVSSKTGRSLRKINSAPTE-----TKSLR	1448
DB	1506	DVSQF---GSFKEDTKMSISEGTVSDK-----SATPVBEGVAEDTVSHMEGVASVS	1553
QY	1449	IASRSTRSHUGLOADVFEVL-----LSPRKRKR	1477

Db 1554 TASVATSSPEPTDSDVSLAEVGSPHSTVEDDLSVSVVQTPTTFTQETEMSGSKEEC 1613
QY 1478 GRKSANNTPEGNPNFR--VIATKSEQSRSVNIASKLSLQES--ESKRRCRKRQSPF 1533
Db 1614 PRPMSISPPDFSPKTAKSRTVPQDHRSEQSSMSIEFG-----QESPEQSLAMDFSRQSPD 1668
QY 1534 PSPVTIG 1540
Db 1669 HPTVGAG 1675

RESULT 43

US-09-538-092-1135
; Sequence 1135, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
; APPLICANT: Mansfield, Traci A.
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538,092
; CURRENT FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CurPatSeqFormatter Version 0.9
; SEQ ID NO 1135
; LENGTH: 2468
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: Polypeptide Accession Number P46821
US-09-538-092-1135

Query Match 3.3%; Score 288.5; DB 4; Length 2468;
Best Local Similarity 19.4%; Pred. No. 1.1e-11;
Matches 304; Conservative 219; Mismatches 547; Indels 497; Gaps 67;

QY 210 PGLTQEALESKKARQLQSF---PEPL-----IIPVLYTSLTHRSRLHETCDIF 259
Db 370 PNIMKRSIEEACFTLQYLKLSMKPEPLFRSVGNITDIPVLFQKMGVGLKLEMYVLNPVK 429
QY 260 AVVKDQRYFVE-----ETVEVIRNNGARLQCTILEVLPPSHQNGFANGVNSV----- 306
Db 430 SKEMQYFMQWGTGNKDKAEFILPNGQEV-----LPISYLTG-----VSSLIWHPA 478
QY 307 -DGETIITISDSDSETQSCSFONGKKDAIDPLLPKYVQP--TKKELH---ESAIVKAT 360
Db 479 NPAEKIIRVLPNGSTQNYILEGLEKHLKD-----FLKQPLATQKDLTGQVPTPVVKQT 533
QY 361 QISRKHLSRDKL-----KULFKOHPCEQPGVVIKIKASSISTYKIAEQDFSYFPF 411
Db 534 KLLQKQAD--SRESLKPAKPLPSKVRKESKEETPEVTKNHVE--KPKVKESKEKVMVK 590
QY 412 DDPPTFIIPANRRGRPKRIHISQEDNVANKOTLASYSKATVE-----RDKLLKQDEM 467
Db 591 DKPVKTETKPSVTEKEVPSKEEPSVFAEVAEKQA--TDVKPKAAKEKTVKKTETKVPKEDK 649
QY 468 KSLAPEKAKLRE---KADALEAKKKEKEDKKEELKKTIVEEERLKKKEKERLKYV- 523
Db 650 KE---EKEKPKKEVAKKEDKTPIKKEKPKKEEVKKEVKEIKKE--EKKGPKKEVKKET 704
QY 524 ---REKERELREERKRYVEYLKQMSKPREDMECDDLKELPE-----PTPVKTRLPPEIFG 576
Db 705 PPKEVKKEVKEEKK---EVKKEEKEPKK-----EIKKLPKDAKSSITPLSEAKXP----- 752
QY 577 DALMVLFLNAGFELFDLQDEFPDGTVLVEALEALVGNDSGECCELLFFFLTAIFQAI 636

Db 753 -----AALKPKVP 760
QY 637 EEEVEAKEQLTADTKGCSLKLSDCTSLSEILRLHILASGADVTSANAKYRYQKRG 696
Db 761 KKEESVKXDSVAAGRPK-----EKGKIVIKKEGAAEAVA-----AAVG 800
QY 697 FDATDDACHMELRLSNPLSVKLSTSVVDLTPGEKMKILHALCGKLLTLVSTRDFIE--- 753
Db 801 TGATTAAYM-----AAAGIAAIGPAKELEAERSLMSSPEDL--TKDFEELKA 845
QY 754 DYVDILROAKOEFRELKAEQHRKEREAAARIRKKEEKLK-----EQEQKM 800
Db 846 BEVDVTIKIQLELIEDEEKLKETEPVEAYVIQKEREVTGPAESPDGIIITTEGEGC 905
QY 801 KEKOEKLKEDQFNSTADISIGEEEREDFTSIESKOTEQEKELDQDMFTTEDDDGSHKR 860
Db 906 EQTPEEL--EPVEKQGVDDIEKFEDSGAGFEESSETGDYEK--AETEEAEEPEDEGEHVC 963
QY 861 GRRGKGQNGKEFTTRQGINCV--TRELLTA-----DEEAL--KQHQRKEKELLBK 910
Db 964 VSASHKSPTEDEESAKAEADAYIREKRESVAGDDRAEEDMDAIEKGEAEQSEEADE 1023
QY 911 IQSAIACNIFPLGRDRMYRVRWIPPSIPGLFIEDY-----SGLTMDM--LLPR 958
Db 1024 DKAEDA-----REEEY-----PEKMEADYVMVVDKAAEAGABEQGFLTT 1067
QY 959 PSSFNQNVQSDPQVSTKTGPELM-----SEBTSNIDQPRDHSVQLPKPVHKPNRWCFY 1013
Db 1068 PTK-QLGAQSPGREPASSIHDETLPGGSESEATASDENRED-----QPEFTAT 1116
QY 1014 SSCQ-----LDQLEALNSRGHRSEALKETLLOEKSRIACAOLARPESEKPHPS 1062
Db 1117 SGYTQSTTIEISSEPTPMDEMSTP-----RDVMSDETNEETESPSQBFVNIT--KYESS 1168
QY 1063 DKPOPDSKPT-----YSRGRSSNAYD-----PSQCAEKQLELRD--- 1099
Db 1169 LYSQEYKPADVTPPLNGFSEGSKTDATDQKYNASASTISPPSSMEEDKFSRSLARDAYC 1228
QY 1100 -----FLLDIEDRIYQGTGAIKVTDHR-----IWSAL--ESGRYELLSBN 1140
Db 1229 SEVKASTITDIKDSI--SAVSSEKVSPSKPSLSPPSPLEKTPLGERSVNFSLTPIEI 1286
QY 1141 KENGLITKTVNEDVEMEID--EOTKVIKORLLGKTETPTSTVSTVASTPQSVSVVHYLA 1199
Db 1287 K-----VSAAEAVAPVSPVPTQEVVEEHCAHCAHCAHCAHCAHCAHCAHCAHCAH 1337
QY 1200 MALFQIEQIEERFLKAPLDASDSGRSVKTVLDRWRESLLSSASLSQVFLHSLTLDREV 1259
Db 1338 -----QSPTDEKSS-----HLPT---EVI 1353
QY 1260 WSKSLNARCKICRKKGDAENNVLCDCDGRGHHTYVVRPKLTVPEGDWFCPECRPKQRC 1319
Db 1354 EKPAVAVPSFFSDAKDENERA-----SVSPMDEPVDSE-----SPIEKV 1394
QY 1320 RRLSFRQPSLESDEVEDSMGDEVDG----- 1349
Db 1395 --LSPLRSPPLIGESAYESFLSADKASRGASPEEKSGKQSPQVSPVSEMTSTS 1452
QY 1350 ---DEEQCQSE-----BEYEVEDQEDDSQBEVEVSLPKRGPQVRLPVKTRGKLSSES 1401
Db 1453 LYQDKQEGSKTDFAPIKEDFQEKKTDDVE-----AMSSQPALALDERKLGDSVPTQI 1505
QY 1402 SRGQOQEGRPYPSQOOSTPKTTVSSKTGRSLRKINSAPTE-----TKSLR 1448
Db 1506 DVQF---GSFKEDTKVSISEGTVSDK-----SATPVDEGVAEDTYSHMEGVASV 1553
QY 1449 IASSTRHSHGLOADVVEL-----LSPRKR 1477
Db 1554 TASVATSSFPPTDDVSPSLHAEVGSHPSTVEDDLSVSVVQTPTTFTQETEMSGSKEEC 1613
QY 1478 GRKSANNTPEGNPNFR--VIATKSEQSRSVNIASKLSLQES--ESKRRCRKRQSPF 1533
Db 1614 PRPMSISPPDFSPKTAKSRTVPQDHRSEQSSMSIEFG-----QESPEQSLAMDFSRQSPD 1668


```

; Sequence 7404, Application US/09949016
; Patent NO. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: C0001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7404
; LENGTH: 2047
; TYPE: PRT
; ORGANISM: Human
; US-09-949-016-7404

```

Query Match	3.3%; Score 281; DB 4; Length 2047;
Best Local Similarity	20.0%; Pred. No. 2.9e-11;
Matches	342; Conservative 268; Mid. Nucleotide 632; Indels 472; Gaps 81
Qy	77 GGGTQPARSAVAAGAAAGACGPERGSPGRRVSPRCLCSGGQVAVGVINGKGR 136
Db	49 GGAKRFLFSERTGFSKLAAMSSWLGIGSGQSGL-----GQVG-GSLASLTGQ 97
Qy	137 RGRDGSRRAPCGREMPLLHRKPFVPPADLRPDDEEVFYCKVTNEIERHVDFFERTIL 196
Db	98 ISNFTKMDLMEGTE-----EVEAEL-PDSRTYKTEAHAILURSENLUK--L 142
Qy	197 CNSL-----VWCAVTRGRLTYQEALESEKKARQ-----NLQSFPEP 234
Db	143 CTDLLEKHEASEIQIKQOSTSYRNQLQOKEVEISHLKARQIALQDQLLKLSAAQSVSG 202
Qy	235 LIIPVLVLT-----ITHRSRLHEICDDI-FAYVKDRYFVEETVEVINNGARLQCTIL 287
Db	203 AGVPATTASSSFAYGISHHPSAFHD--DDMDFG--DIISQOEINRLSNEVSRLESEV- 256
Qy	288 EVLPFPHQNGFANGVNSVDEGTIIISDSDS-ETQSCSPONGKKDAIDPLFYKYQVP 346
Db	257 -----GHWRIH-AQTSKAQGTDSQOSEICKLQNIKE-----LKNRSQE 296
Qy	347 TKKELHESAIV-----KATOISRRKHLFSRDLKLFLKQCEPOEGVIKIKASSLSTYK 400
Db	297 IDHOHEMSVLQNAHQKLTEISRR-----HREELSDY-EERIELEULNQGGSG- 346
Qy	401 IAEQDFSYFFPDPPPTFFSPANRRRRGPPRIHISQEDNVANQTLLASYSRK----- 453
Db	347 VIETDLSKIY-----EMQKTIQVLOLEKVESTKQMEQLEDKIDINKK 389
Qy	454 ----ATKERDKLLQ-----BEMKSLAFEKAKLKR-----EKADALEAKK- 490
Db	390 LSSAENDRDLIRREQEQQLNVEKQIMBECENLKLECSKLQPSAVKQSDTMTKEKERILAQ 449
Qy	491 -----EKEDKEKKEELKKIVVEERLKKKEEKERLKVZ-REKEREKLBEKKRY 538
Db	450 ASVEEVFRLQOALSDAENEIMRLSSLNQDNSLAEDNLKLNRIEYLEKEKSLLSQEK- 506
Qy	539 VEYLKQWSKPREDMCEDDLKELPPTPVKTRLPPEIFGDLMLWLEFLNAGFELPDQDEF 598
Db	507 -----EEIQMSLLKLNNEFYVTKSTATRDLSLDS-----ELHDLR- 541
Qy	599 PDGVTLLEVLBEALVGNDSQGLPCELLFFFLTAIFAQIAAEHEEE-----VAKEQLTDADT 652
Db	542 ---LNLEAKQELNQSISEKET-----LIATIEELDQNOEQATKHWILLIKDQLSKQON 591
Qy	653 KGCSLKS- --LDLDSCTLSEILRLHIILASGADVTSANAKRYQKRGGPDATDDACMEURL 709

592	Db	EGDSIISKLKQDLN----	DBKKRVHOLE-----	DDKMIDITKELDV	627
710	Qy	SNPSLVKLLSSTSYDLTPGKMKILHALCGKLLTLVSTDFIEDYVDILROAKQ----	764		
628	Db	QKEKLIQ--SEVALND-----	LH-----	LTQKLEDKVENLDQNLKQSNVS	670
765	Qy	EFRELKABQHRKEREEMAAIRK-----	KEEKLKEQOKMEKOEKLKEDE	811	
671	Db	QKENLELKEHIRQNEEELSIRNELMOSLNQDSNFKDTLLKEREAEVNLKONLSELE	730		
812	Qy	QRNSTA-----DISIGEEER----	EDF-----	DTSIESKDEKOKELD	844
731	Db	QLNENLKVAFVDMENKVLVACEDVRHQLESCLAGNQLSLEKTVIVETLKWEGEIE	790		
845	Qy	QDMFTDEDDPGSHKRRGRKRGONGKEFTROEQINCVTRELTLTAADEEALKOEHORKE	904		
791	Db	AELCW-----AKRLLLEEANKYEKTEIELSARNLNTSALQ----	EHEHLIKLNQKOD	840	
905	Qy	KELLEKIOSAIACTNLFPGLGRDRMYRYMFPIS-----	PGFLIE----	EDYSG	949
841	Db	METAEULK-----NIEQMDTDHKETKDVLSLSLEEQKQLTQLINKEKPIEKLKERSSK	894		
950	Qy	LTEDM-----LLPRPSSFQNNVQSDPPQVSTKTGE--	PLMSESTSNIDGPRDHSVOLPK	1002	
895	Db	LOEELDKYSQALAKNEILRTQIEEKORSGLSMEENHLQEELERUEQSRAPVADPK	954		
1003	Qy	PVKHPNRWCYSCBOLDQJLIEALNS--RGHRESALK--	ETLLOEKSRICQALARFEEKF	1059	
955	Db	TL-----DSVTELASEVSQNLTKHELEBEEIKHHQKIIIDONQSKMLQLSQEOQ--	1004		
1060	Qy	HFSDKPOPDSPKPTYSRGRSSNAYDPQOMCA--EKQELRLRDLFLDIEDRIYQCTGLGAIKV	1118		
1005	Db	-----KKEMD-----	EFRYQHQMATHQTLQFLE-----	KDEISIKSLQXTI	1044
1119	Qy	TDRIHWRSALESGRYELLGSE-----	NKENGIIKTVYNEDVEEMEIDEQTKVIVKDRLLGI	1173	
1045	Db	TOLHEERODIOTNSDIFQETKVQOSLNIENGSEK--	HDLSSKAETERLVKGI--KERELEI	1100	
1174	Qy	KTETPSTVSNASTPQSVSSVHYLAWALFOIBQGIERRFLKAPLADSDSGRSYKTVLDR	1233		
1101	Db	KLANEKXISUTKIQDLSKDXDEVGKLTQIIQQKDLQIQ-----	ALHARISSTSTHTQDVVY	1154	
1234	Qy	WRSELSSA--SLSQVFLHLSTLDRSVIWSKSILNARCKTCKKGDAAENMVLDCGCDRGHH	1292		
1155	Db	LOQOLQAYAMEREKVFVNLNEKTRENSHLKTEYHKMDIIVAAK--EAALIKLQD-----	1206		
1293	Qy	TYCVRPKLTVPGE--DNFPCPCRPQRCRLUSFRQPSLESDVE--DSMGGE-----	1343		
1207	Db	-----ENKLSLTPRESSQDMF-----	RETIQNLRSIITREKIDIEIDALSQKQCTLL	1252	
1344	Qy	-----DDEVGDDEBEGQEEYEVEQDDDSOBEEEVSLPKRGPPQVRLPVKTRGKLSSSF	1400		
1253	Db	AVLQTSSTGNEAGVNSNGFEELLQERD--	KLQVKKMEEWKQVMTVV-----	1300	
1401	Qy	SSRGQOQEGRPYSRQSQSTPKTTVSSKTGRSLR-----	KINSAPPTETKSLRIASR--S	1453	
1301	Db	-----QNMCHESAQLQEELHQLQAQVLVDSDNNSKLQVDYTGLIQSYEQNETKLKMFQBELAQ	1358		
1454	Qy	TRISHGPL--QADV-----	VELLSPRKRRGRKSKANTPENSFPNFPRIATKSSQSR	1508	
1359	Db	VQHSIGOLCNCKDILGKLDIITSPOUS-----	SASLLTPOSAE-----	CLRASKEVLS	1409
1509	Qy	VNIASKLSLQESBKRCRKRQSPSPVTLGRRSSRGQGVHLSAFEOLVVELVR--H	1566		
1410	Db	-----SELLQOELELR--KSLQEKDATIRTL-----	QENNRHRLSDSTAATSELERKEH	1456	
1567	Qy	DDSWPFLKLVSKTQVPDYVDIILKPPALNIIREK	1600		
1457	Db	EOTDSEIK-----OLKEKQDVLOK-----	LLKEK	1480	

RESULT 46
US-08-353-700-1

```

; Sequence 1, Application US/08353700
; Patent No. 5599919
; GENERAL INFORMATION:
; APPLICANT: YEN, TIMOTHY J.
; APPLICANT: RATNER, JEROME B.
; TITLE OF INVENTION: NUCLEIC ACID ENCODING A
; TITLE OF INVENTION: TRANSIENTLY-EXPRESSED KINETOCHORE PROTEIN,
; TITLE OF INVENTION: AND METHODS OF USE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DANN, DORFMAN, HERRELL AND SKILLMAN
; STREET: 1601 MARKET STREET, SUITE 720
; CITY: PHILADELPHIA
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/353.700
; FILING DATE: 09-DEC-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: REED, JANET E.
; REGISTRATION NUMBER: 36,252
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 563-4100
; TELEFAX: (215) 563-4044
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3248 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: HUMAN
; US-08-353-700-1

```

Query Match 3.2%; Score 279; DB 1; Length 3248;

Best Local Similarity 19.7%; Pred. No. 8.2e-11;

Matches 326; Conservative 240; Mismatches 580; Indels 506; Gaps 78;

```

Qy 243 TSLTHSRSLRHLICDDIFAYVKD-RYFVETVEVIRNNGARLQCTILEVLPSPHQGFANG 301
Db 1778 TERTPKGDVHQID-----KDAQDMLDIEKITETGA-----VKPTGECGSGQSP 1823
Qy 302 HVN-SVDGETIISDSDSETQSCSFQNGKKDAIDPLLF-----KYKVQPTKKE 350
Db 1824 DTNYEPPGE-----DKTQSGSECISELFSFGPNALVPMDFLGNQEDIHNLQLRVKETSNE 1878
Qy 351 ----LH-----ESAIVKATQISRKHLPFSRDKLFLK-QHC-----EPQEGVIKKA 393
Db 1879 NLRLLHVIEDRDRKVESLNNKEMKLDKHLH---QEVQMTKIEACIELEKIVGELKKEN 1935
Qy 394 SLSLTYKIAEQDFSYFFDDPTFTFSPANRRRGRPPKRIHSQEDNVANKQTILASYSRK 453
Db 1936 SDSL-----EKLEYFSCDHQELL-----QRVETSEGLN-----SDLEMHADR 1972
Qy 454 ATKE-----RDKLLKQE-EMKSLAFKAKLKR-----KADALBAKKKEKEDK 495
Db 1973 SSREDIGNVAKVNDWKRFLDVENLSRIRSEKASIEHEALYLEAD-LEVVQTEKLCL 2031
Qy 496 EKKREELKKIV---EEERLKKKEKERLKVREKEREK-----LREKRKYVYILKQWS 546
Db 2032 EKDNENKQKIVLCLEELLSVVTSERNQURGLDWTMSKTTALDQLSEKMKKEKTOELE--S 2089
Qy 547 KPREDMECDLDELPEPTPVKTRLPPEIFGDALMVLEFLNAGFELFDLQDEFDPDGVTLVEV 606

```

```

Db 2090 HQSECLHCIOVAE--ABVKEKTELLQTLSSDVSELLK-----DKTHLQEK-----LQS 2135
Qy 607 LEEALVGNDSG-----PLCELLFPFLTAIFAIAAEEVEEVAKEQLTDTADTKCSLSKSLDL 663
Db 2136 LEK-----DSQALSLTKCEL-----ENQIAQ-LNKEKELLVKE-----SESQARLSESDYE 2181
Qy 664 SCTLSEILRLHLILASGADVTSANAKYRYQKRGGPDATDDACMEURLNPNLSLVKLSSTSV 723
Db 2182 KLVSKALEALVEKGE-----FAURLS-----TQEEV 2210
Qy 724 YDLTPG-EKMKI-----LHALCGKLLTLVSTRDFIEDYVDILR-----QAKQFR 767
Db 2211 HQLRGGIEKLVRIEADEKKQLH-IAEKLERERENDSLKDKVENLSRELQMSEENQELV 2269
Qy 768 ELKAEQHRKERE-----EAAAR-----IRKREKLEKQKQKKEKQE-----805
Db 2270 ILDAENSKAEVETLKTQIEEMARSLKIFELDLVTLRSEKENLTQIOEKQQLSELDKLL 2329
Qy 806 -----KLKEDBQRNSTADISIGREEREDFDTISIESKOTEKEL-----DQDMF--T 849
Db 2330 SSFKSLLEKEQ-----AEIQKEEK-----TAVEMLQNLKELNEAVALCGDGEIMKAT 2381
Qy 850 EDEDDPGSHKRG--RRGKRGQNGFKFTROEQINCVTREL-----887
Db 2382 EQSLDPIEBEHQLRNSIEKLRLARLEADEKKQL-CVLQQLKESSEHADLLKGRVENLERE 2440
Qy 888 -----LTADEBEALKQEHORKEKELLE-KIOSAIACTNIFPLGRDMYRRYWFISPLGLF 942
Db 2441 LEIARTNQEHAALAEANSKGEVETLKAKIEG-----2471
Qy 943 IEEDYSGLTEDMLLPR--PSSFQNNVQSDPOVSTKTGTEPLMSESTNIDQGRDHSVOL 1000
Db 2472 MTQSLRGLDLDVTVIRSEKENLTNLEKQERISEL---EIINSSFNILQEKQEKQVQM 2528
Qy 1001 PK-----PVHKPNRWC-----FYSCQOLD-QLIEALNSR 1029
Db 2529 KEKSTAMEMLQTLKELNERRVAALHNDQBACKAQNLSSQVECLEKAQLQLGLDEA 2588
Qy 1030 GHRESALKET---LLQE-----KSRIQAQLARFSEKHFSDKPPQDSKPTYSRRSSNAY 1082
Db 2589 KNNYIVLQSSVKGLIQEVEDGKQLEKQEBEISLRKNQIQDQEQLVSKLSQVEGE-----2643
Qy 1083 DPSQWCAEKOELRLRDLFDLDIEDRI-----YQGTGAIKVTRDHTWRSALESGRY 1133
Db 2644 --HOLWKEQNLN---LRNLTVLEEQIOVLQSKNASLODTLEVLOSXKNL-ENLELTKM 2698
Qy 1134 ELIASEENKENGIIKTWNEDVEEMEIDRQTKVIVKDRLLGIKTETPTSTVNASTPQSVSS 1193
Db 2699 DKMSFVEKVNKWTAKETELQREMHMAQKTAELQELSGEKNRLAGELQLLLEBKSKD 2758
Qy 1194 VVHYLAWALFOIEGIERFLKAPLDAS-----1221
Db 2759 QLKELT-----LENSELKSLDCMHWKQDQVEKGVREIEAAYQLRLHEABKKHQAL 2809
Qy 1222 --DSGRSKYTVLDNRWRESLJSSAS-LSQVFLHLSTLDRSVTWSKSIINARCKICR--KKG 1276
Db 2810 LLDNKOYEVEIQYREKLTSEKECLSQLEIDLKSSKEELNNLSKATQIILEELKKT 2869
Qy 1277 DAENNVLCGDCDRGHHTYCVRPKLTVPEDGDFWPECRCRQRRLSPFRQPSLESBDV 1336
Db 2870 KMDNLKYNVQLKKENER--AQGKMKLIKS---CKQLEEEKEIILQKELSQLQAQEKQKT 2924
Qy 1337 EDSMGDEDDVDGDEBEGQEEEEEYVEQDE-----DDSQEEVEVSL---1378
Db 2925 GTVMDTKVDELTTIKELKETLEKTEADEYLDKYGSLLSISHEKLEKLEKEMLETQVAHL 2984
Qy 1379 -----PKGRPOV-----RLPVKTRGKLSSEFS-SRQGOQ-----EPGRYPSRS 1416
Db 2985 CSQOSKQDSRGSPLLGVPVPGSPFIPSVTEKRLSSGQNKASGKQRSGGIWENRGPP---3041
Qy 1417 QQSTPKTIVTSKTKGRSLRKINSAPPTETKSLRIASRSTRSHSGHGFLOADVVFVLLSPRRKR 1476

```


Db 3042 ---TPAT-----PSSFKSKKAVMSGIHPAEDTEGTGTEPPEGLPEVVK 3082
Qy 1477 RGRKANNTPENPNFPRVIATKSSBQSRSVNIA-SKLSIQE---SESKR-----RC 1526
Db 3083 KGFADI-PTGKTSPIYLRTTMAITTSPLRLAAQKLALSPLSLGKENLAESSKPTAGGSR 3141
Qy 1527 RKRQSPSPV---TLGRSSGROGGVHLSA-----PEQ 1558
Db 3142 QKVVAQRSPVDSGTHILREPTTKSPVNNLPERSPTDSPREGRLVKRGLRVPAPKLDWSQ 3201
Qy 1559 LVVELVRH-----DSSWPFKLKSVKIOVPDY 1584
Db 3202 LAVTRSRSEALCVSDPWEVQSLDRCLQDF 3233

RESULT 47
PCT-US95-16216-1
; Sequence 1, Application PC/TUS9516216
; GENERAL INFORMATION:
; APPLICANT: Yen, Timothy J.
; APPLICANT: Rattner, Jerome B.
; TITLE OF INVENTION: Nucleic Acid Encoding a Transiently
; TITLE OF INVENTION: Expressed Kinetochores Protein, and Methods of Use
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESS: Dann, Dorfman, Herrell and Skillman
; STREET: 1601 Market Street Suite 720
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/16216
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/353,700
; FILING DATE: 09-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Reed, Janet E.
; REGISTRATION NUMBER: 36,252
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 563-4100
; TELEFAX: (215) 563-4044
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3248 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
PCT-US95-16216-1

Query Match 3.2%; Score 279; DB 5; Length 3248;
Best Local Similarity 19.7%; Pred. No. 8.2e-11;
Matches 326; Conservative 240; Mismatches 580; Indels 506; Gaps 78;

Qy 243 TSLTHRSRLHEICDDIFAYVD-RYFVEETVEVIRNNGARLOCTILEVLPSPHQNGFANG 301
Db 1778 TERTPKHDVHOICD-----KDAQOQDLNLDIEKITETGA-----VKPTGECSGEQSP 1823
Qy 302 HWN-SVDCGETIISDDSETQSCSFQNGKKDAIDPLLF-----KYKVQPTKKE 350
Db 1824 DTNYEPPGE-----DKTQGSSECSIELSFGPNALVPMDFLGNQBDIHNLQURVKVETNE 1878
Qy 351 ----LH-----ESAIVKATQISRRKHLFSRDKLFLK-QHC-----EPQEGVIKKA 393

Db 1879 NLRHLHVIEDRRDKVESLNLNEMKELDSKLHL---QEVQMTKIEACIELEKIVGELKKEEN 1935
Qy 394 SSLSTYKIAEQDFSYFFPDPTTFIFSPANRRRGPPKRIHISOEDNVANQOTLASYSK 453
Db 1936 SDLS-----EKLYFSCDHQELL-----QRVETSEGLN-----SDLEMHADK 1972
Qy 454 ATKE-----RDLLKQF-EMKSLAFKAKLKRE-----KADALBAKKKEKEDK 495
Db 1973 SSREDIGNVAKVNDMSKERFLDVENELSRIRSEKASIEHEALYLEAD-LEVVOETKLC 2031
Qy 496 EKKREELKIV---EERLKKKEKERLKVREKERK-----LREKKRYVVELKOWS 546
Db 2032 EKDNENKQKIVCLEEELSVVTSERNLQGLDTSKKTALDQLSQKKKKTQELB--S 2089
Qy 547 KPREDMECDLDELPEPTPVKTRLPPEIFGDMVLEFLNAGFELFDLQDFPDGVITLEV 606
Db 2090 HQSECLHICIVAE--AEVKEKTELLOTLSSDVSELK-----DKTHLOEK-----LQS 2135
Qy 607 LEEALVGNDSRG---PLCELLFFFLTAIFAABEEEEVAKEQLTADTKGCSLKSLDL 663
Db 2136 LEK-----DSQALSITKCEL---ENQIAQ-LNKEKELLVKE---SESQARLESSE 2181
Qy 664 SCTLSEILRLHILASGADVTSANAKYRQKGGFDATDDACWELRLSNPSLVKLSSTSV 723
Db 2182 KLVSKALEAALVKEG-----LHALCGKLLTVSTRDPIDYVDILR-----QAKQF 767
Qy 724 YDLTPG-EKMKI-----IAEKLKERENDSLKOKVENLERELQWSENQELV 2269
Db 2211 HOLRRGIEKLVRIEADKKQLH-IAEKLKERENDSLKOKVENLERELQWSENQELV 2269
Qy 768 ELKAEQHRKERE-----EAAAR-----IRKREKELKEQEQMKKEQ- 805
Db 2270 ILDAENSKAEVETLTKTOIEEMARSLKTFELDLVTLRSEKENTKIQIEQOGLSBLDK 2329
Qy 806 -----KLKEDQORNSTADISIGEEREDFDTSIESKOTEQEL-----DQDMF--T 849
Db 2330 SSFKSLLEEKG---AEIQIEESK-----TAVEMLQNLKELNEAVALCGDQEI 2381
Qy 850 EDEDDPGSHKRG---RRGKRGQNGFKFTROQINCVTREL----- 887
Db 2382 EQSLDPIFEEHQJRNSEIKLARLEADEKKQL-CVLQOLKESEHADLLKGRVENLRE 2440
Qy 888 ----LTADDEEALKQEHORKEKELLE-KIQSAIACTNIFPLGRDRMYRYYWIFPSIG 942
Db 2441 LEIARTQEHAALEAENSKGSEVETLKAKIEG----- 2471
Qy 943 IEEDYSGLTEDMLPR--PSSFQNNVQSDQVSTKTGCEPLMSESTNIDQPRDHSVOL 1000
Db 2472 MTQSLRGLELDVVTIRSEKENLTNELOKEQERISEL---EINSFENILOKEQEKV 2528
Qy 1001 PK-----PVHKNRWC-----FVSCCEQLD--OLTEALNSR 1029
Db 2529 KEKSTAMEMLQTLKELNERVAALHNDQEAACKAEQNLQSVCELEKQLGLQLEA 2588
Qy 1030 GHRESALKET---LLQE---KSRICAQLARFSEKHFSDKPPQDSKPTYSRGRSSNAY 1082
Db 2589 KNNYIVLQSSVKGLIQEVEDGKQLEKDEISLKNQIQDQEQLVSKLSQVEGE----- 2643
Qy 1083 DPSQMCAPKQLELRDLFLDIEDRI-----YQOTLGAIKVTDRIHRSALLESRY 1133
Db 2644 --HQLWKEQNLE--LRNLTVLEQKIQVLQSKNASLQDTLEVLOQSSYKNL-ENELELTKM 2698
Qy 1134 ELLSEENKENGIIKTVNEDEVEEIDEOTKVIKDRLLGIKTETPTSTVSTNASTPQSVSS 1193
Db 2699 DKMSFVEKVNQWAKETELQREHMAQKTAELQESLSEKRNRLAGELQLLEELKSKD 2758
Qy 1194 VVHYLAMALFQIEOGIERFLKAPLDAS----- 1221
Db 2759 QLKELT-----LENSELKSLDCMHKQDOVEKGVREIAEYQRLHAEAKHQAL 2809
Qy 1222 --DSGRSYKTVLDRWRSELLSSAS--LSQVFLHSTLDRSVTWSKILNARCKICR--KKG 1276

Db 2810 LLDTNKQVEIQTREKLTKEECLSSQKLEIDLKSSKEELNLSKATQIILELKKT 2869
Qy 1277 DAENMVLCDGCDRGHHTYCVRPKLTVPBGDFCPCRPKQRCRLSPRPSLESDEDV 1336
Db 2870 KMDNLKYNQJLKKENER--AQCKMKLLTKS---CKQLEEKKEIQLKELSQJQAQEKQRT 2924
Qy 1337 EDSMGEDVDGDDEEQSQSEEEVEVBQDE-----DSDQEEEEEVSL--- 1378
Db 2925 GTVMDTKVDELTTTELKELKELEKTEKADDEVLDKYCSLLISHEKLEKAKEMLETOVAHL 2984
Qy 1379 -----PKRGRPOV-----RLPVKTRGKLSSEFS-SRQOQO-----EPGRYPERS 1416
Db 2985 CSQSQKQDSRSPGLLPVPGPSPITPSVTEKRLSSQNKASGKQRSSGIWENGKGP--- 3041
Qy 1417 QOSTPKTTVSSKTGSLRKINSAPTETKSLRIASRSTRHSHGPLQADVVELLSPRKR 1476
Db 3042 ---TPAT-----PESFSKSKKAVMSGIHPAEDTEGTEFEPEGLPEVVK 3082
Qy 1477 RGRKSANNTSPNSPFPNFRVIATKSSQSRSVNIA-SKLSLQE---SESKR-----RC 1526
Db 3083 KGFADI-PTGKTSPIYLRTTMTATSPRLAAQKLSPLSLGKENLAESSKPTAGGSR 3141
Qy 1527 RKQSPSPSPV---TLGRRSSGRQGVHLSA-----FEQ 1558
Db 3142 QKVKAQRSPVDSGTILREPTTKSVVNNLPERSPTDSPREGLRVKRGLVPAPKLDWSQ 3201
Qy 1559 LVVELVRH-----DSDWPFLLKLSKIQVPDY 1584
Db 3202 LAVRTSRSEALCVSDPWEVQSLDRCLQDP 3233

RESULT 48
US-09-418-710-43
; Sequence 43, Application US/09418710
; Patent No. 6596482
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042001
; CURRENT APPLICATION NUMBER: US/09/418,710
; CURRENT FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; SOFTWARE: FastSEQ for Windows Version 4.0
; NUMBER OF SEQ ID NOS: 73
; SEQ ID NO 43
; LENGTH: 45
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-418-710-43

Query Match 3.2%; Score 276; DB 4; Length 45;
Best Local Similarity 100.0%; Pred. No. 2.2e-13;
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1269 CKICRKGDAENMVLCDGCDRGHHTYCVRPKLTVPBGDWFPCPEC 1313
Db 1 CKICRKGDAENMVLCDGCDRGHHTYCVRPKLTVPBGDWFPCPEC 45

RESULT 49
US-09-418-710-56
; Sequence 56, Application US/09418710
; Patent No. 6596482
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042001
; CURRENT APPLICATION NUMBER: US/09/418,710

; CURRENT FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 56
; LENGTH: 45
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-418-710-56

Query Match 3.2%; Score 276; DB 4; Length 45;
Best Local Similarity 100.0%; Pred. No. 2.2e-13;
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1269 CKICRKGDAENMVLCDGCDRGHHTYCVRPKLTVPBGDWFPCPEC 1313
Db 1 CKICRKGDAENMVLCDGCDRGHHTYCVRPKLTVPBGDWFPCPEC 45

RESULT 50
US-09-839-479-42
; Sequence 42, Application US/09839479
; Patent No. 6727222
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042002
; CURRENT APPLICATION NUMBER: US/09/839,479
; CURRENT FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 42
; LENGTH: 45
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-839-479-42

Query Match 3.2%; Score 276; DB 4; Length 45;
Best Local Similarity 100.0%; Pred. No. 2.2e-13;
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1269 CKICRKGDAENMVLCDGCDRGHHTYCVRPKLTVPBGDWFPCPEC 1313
Db 1 CKICRKGDAENMVLCDGCDRGHHTYCVRPKLTVPBGDWFPCPEC 45

Search completed: March 9, 2005, 14:56:25
Job time : 95.454 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 9, 2005, 13:18:56 ; Search time 144.658 Seconds
(without alignments)
5405.490 Million cell updates/sec

Title: US-10-702-148-27
Perfect score: 7967
Sequence: 1 MAPLGRKPFPLVPLPGBE.....NCFMMLVNTQFCMALTDVT 1527

Scoring table: BLOSUM62
Gap 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 65 summaries

Database : Uniprot 03: *
1: uniprot_sprot: *
2: uniprot_trembl: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7674	96.3	1483	1 BAIH_HUMAN	Q9u190 homo sapien
2	6989	87.7	1479	1 BAIH_MOUSE	Q9u277 mus musculus
3	4940	62.0	955	2 Q86UJ6	Q86u16 homo sapien
4	4156	52.2	808	2 Q6PLK4	Q6plk4 homo sapien
5	3740	46.9	1079	2 Q8UVR4	Q8uvr4 xenopus lae
6	3029.5	38.0	657	2 Q8CAU9	Q8cau9 mus musculus
7	2878.5	36.1	777	2 Q6Q006	Q6gq06 xenopus lae
8	915.5	11.5	1556	1 BAIH_HUMAN	Q9v9t4 h bromodoma
9	664	8.3	1476	2 Q9V9T4	Q9v9t4 drosophila
10	662	8.3	1476	2 Q9NG24	Q9ng24 drosophila
11	661	8.3	1476	2 Q9Y0W1	Q9y0w1 drosophila
12	657.5	8.3	1427	2 Q9N5L9	Q9n5l9 caenorhabdi
13	648	8.1	1357	2 Q7PRP9	Q7prp9 anopheles g
14	640.5	8.0	892	2 Q6P1D9	Q6pld9 mus musculus
15	553	6.9	1698	2 Q6YI94	Q6vi94 xenopus lae
16	537.5	6.7	1905	2 Q68D18	Q68d18 homo sapien
17	536	6.7	1372	1 BAIH_HUMAN	Q9u1f8 homo sapien
18	534.5	6.7	1878	1 BAIH_MOUSE	Q9u1f5 homo sapien
19	518.5	6.5	1850	1 BAZA_MOUSE	Q91ye5 mus musculus
20	507	6.4	718	2 Q8C8D1	Q8c8d1 mus musculus
21	502	6.3	2130	1 BAIH_CHICK	Q9del3 gallus gall
22	497	6.2	1202	2 Q7QBK2	Q7qbk2 anopheles g
23	489	6.1	2060	2 Q7Q3S9	Q7q3s9 anopheles g
24	455.5	5.7	1370	2 Q45075	Q45075 caenorhabdi
25	448.5	5.6	1376	2 Q23590	Q23590 caenorhabdi
26	407	5.1	627	1 BAIH_XENLA	Q8uvr5 xenopus lae
27	370.5	4.7	1508	2 Q75JF5	Q75jp5 dictyosteli
28	363.5	4.6	10578	2 Q8ISF5	Q8isf5 caenorhabdi
29	363.5	4.6	18519	2 Q8ISF6	Q8isf6 caenorhabdi
30	363.5	4.6	18534	2 Q8ISF7	Q8isf7 caenorhabdi
31	354.5	4.4	7210	2 Q9V7G8	Q9v7g8 drosophila

32	354.5	4.4	9270	2 Q8MLD9	Q8mld9 drosophila
33	353	4.4	886	2 Q8OU42	Q8ou42 mus musculus
34	347.5	4.4	1264	1 ITCL YEAST	P53125 saccharomyc
35	344	4.3	2759	2 Q9VID9	Q9vid9 drosophila
36	343	4.3	3080	2 Q9V602	Q9v602 drosophila
37	337.5	4.2	1837	2 Q75PR8	Q75pr8 hemiceutrot
38	332.5	4.2	3109	2 Q9BMQ0	Q9bmq0 drosophila
39	329	4.1	1230	2 Q6BV90	Q6bv90 debaryomyce
40	325	4.1	1790	1 US01 YEAST	P25386 saccharomyc
41	324.5	4.1	969	2 Q6CAZ9	Q6caz9 varrowia li
42	324	4.1	617	2 Q6Z178	Q6z178 oryza sativ
43	324	4.1	1268	2 Q7LGT1	Q7lgt1 saccharomyc
44	324	4.1	1790	2 Q07380	Q07380 saccharomyc
45	318.5	4.0	1980	2 Q6FWE0	Q6fwe0 candida gla
46	317	4.0	1661	2 Q06166	Q06166 plasmodium
47	315.5	4.0	2042	2 Q6BUQ9	Q6buq9 debaryomyce
48	309.5	3.9	1510	2 Q25920	Q25920 plasmodium
49	309	3.9	1596	2 Q81J44	Q81j44 plasmodium
50	306	3.8	3616	2 Q9W6V0	Q9w6v0 gallus gall
51	305.5	3.8	1864	2 Q6P5H2	Q6p5h2 mus musculus
52	304.5	3.8	1972	2 Q8R384	Q8r384 mus musculus
53	304.5	3.8	1984	2 Q692X3	Q692x3 mus musculus
54	303	3.8	3259	1 G0B1_HUMAN	Q14789 homo sapien
55	302.5	3.8	1345	2 Q9Y2L2	Q9y2l2 homo sapien
56	301.5	3.8	683	2 Q960I3	Q960i3 drosophila
57	301.5	3.8	1431	2 Q96T23	Q96t23 homo sapien
58	300.5	3.8	1972	1 MYHB_MOUSE	Q08638 mus musculus
59	298.5	3.7	2116	1 MYS2_DICDI	P08799 dictyosteli
60	298	3.7	1434	2 Q81492	Q81492 plasmodium
61	298	3.7	2492	1 ATRX_PONPY	Q7Ygm3 pongo pygma
62	297.5	3.7	1927	2 Q75WX9	Q75wx9 helicobacte
63	297	3.7	3111	2 Q9VH10	Q9vh10 drosophila
64	296	3.7	1821	2 Q9R0C4	Q9r0c4 mus musculus
65	296	3.7	2492	1 ATRX_HUMAN	P46100 homo sapien

ALIGNMENTS

RESULT 1	BAIB_HUMAN	STANDARD;	PRT; 1483 AA.
ID	Q9UIG0; Q95039; Q95247; Q95277;		
AC	DT 28-FEB-2003 (Rel. 41, Created)		
DT	DT 28-FEB-2003 (Rel. 41, Last sequence update)		
DT	DT 25-OCT-2004 (Rel. 45, Last annotation update)		
DE	Bromodomain adjacent to zinc finger domain protein 1B (Williams-Beuren syndrome chromosome region 9 protein) (WBRS9) (Williams syndrome transcription factor) (HWALP2).		
GN	Name=BAZ1B; Synonyms=WBSC10, WBSCR9, WSTF;		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1] SEQUENCE FROM N.A. (ISOFORM 1).		
RP	MEDLINE=99077764; PubMed=9858827;		
RX	Peoples R.J., Cisco M.J., Kaplan P., Francke U.;		
RA	"Identification of the WBSCR9 gene, encoding a novel transcriptional regulator, in the Williams-Beuren syndrome deletion at 7q11.23.";		
RT	Cytogenet. Cell Genet. 82:238-246(1998).		
RL	[2] SEQUENCE FROM N.A. (ISOFORM 1).		
RN	MEDLINE=99047530; PubMed=9828126; DOI=10.1006/geno.1998.5578;		
RX	Lu X., Meng X., Morris C.A., Keating M.T.;		
RA	"A novel human gene, WSTF, is deleted in Williams Syndrome.";		
RL	Genomics 54:241-249(1998).		
RN	[3] SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).		
RP	TISSUE-Testis;		
RC	MEDLINE=20130112; PubMed=10662543; DOI=10.1006/geno.1999.6071;		
RX	Jones M.H., Hamana N., Nezu J., Shimane M.;		
RA	"A novel family of bromodomain genes.";		
RT			

Genomics 63:40-45(2000).
 [4]
 RN SEQUENCE FROM N.A.
 RA Minx P., Graves T., Duckels G., Harrison M., Waterston R.;
 RL Submitted (DEC-1999) to the EMBL/GenBank/DDBJ databases.
 [5]
 RN FUNCTION.
 RP MEDLINE=21977304; PubMed=11980720; DOI=10.1093/emboj/21.9.2231;
 RX Bozhonok L., Wade P.A., Varga-Weisz P.;
 RA "WSZF-1SWI chromatin remodeling complex targets heterochromatic
 RT replication foci.";
 RL EMBO J. 21:2231-2241(2002).
 CC -!- FUNCTION: Forms a chromatin remodeling complex that mobilizes
 CC nucleosomes and reconfigures irregular chromatin to a regular
 CC nucleosomal array structure.
 CC -!- SUBUNIT: Interacts with ISWI (imitation SWI protein) to form the
 CC WSZF-1SWI chromatin remodeling complex (WICH).
 CC -!- SUBCELLULAR LOCATION: Nuclear (Potential). Accumulates in
 CC pericentromeric heterochromatin during replication.
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=1;
 CC IsoId=Q9UIG0-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=Q9UIG0-2; Sequence=VSP 000552;
 CC -!- TISSUE SPECIFICITY: Ubiquitously expressed with high levels of
 CC expression in heart, brain, placenta, skeletal muscle and ovary.
 CC -!- DEVELOPMENTAL STAGE: Expressed at equal levels in 19-23 weeks old
 CC fetal tissues.
 CC -!- DISEASE: Haploinsufficiency of BAZ1B may be the cause of certain
 CC cardiovascular and musculo-skeletal abnormalities observed in
 CC Williams-Beuren syndrome (WBS), a rare developmental disorder. It
 CC is a contiguous gene deletion syndrome involving genes from
 CC chromosome band 7q11.23.
 CC -!- SIMILARITY: Belongs to the WAF family.
 CC -!- SIMILARITY: Contains 1 bromodomain.
 CC -!- SIMILARITY: Contains 1 DDT domain.
 CC -!- SIMILARITY: Contains 1 PHD-type zinc finger.
 CC -!- SIMILARITY: Contains 1 WAC domain.
 CC -!- CAUTION: Ref.2 sequence differs from that shown due to frameshifts
 CC in positions 1031, 1042 and 1422.
 CC -!- CAUTION: Ref.3 sequence differs from that shown due to a
 CC frameshift in position 1478.
 CC -!- CAUTION: Ref.4 sequence differs from that shown due to erroneous
 CC gene model prediction.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AF084479; RAD08675.1; -;
 CC EMBL; AF072810; AAC97879.1; ALT FRAME.
 CC EMBL; AB032253; BAA99210.1; ALT FRAME.
 CC EMBL; AC005074; RAD04720.1; ALT SEQ.
 CC EMBL; AC005089; -; NOT ANNOTATED_CDS.
 CC PDB; 1P62; NMR: A=1185-1235.
 CC TRANSFAC; T04145; -;
 CC Genew; HGNC:961; BAZ1B.
 CC MIM; 605681; -;
 CC GO; GO:0005634; C:nucleus; NAS.
 CC GO; GO:0003700; F:transcription factor activity; NAS.
 CC GO; GO:0008270; F:zinc ion binding; NAS.
 CC GO; GO:0006350; P:transcription; NAS.
 CC InterPro; IPR001487; Bromodomain.
 CC InterPro; IPR004022; DDT.
 CC InterPro; IPR001965; Znf PHD.
 CC Pfam; PF00439; Bromodomain; 1.
 CC Pfam; PF06628; PHD; 1.
 CC PRINTS; PR00503; BROMODOMAIN.

DR SMART; SM00297; BROMO; 1.
 DR SMART; SM00571; DDT; 1.
 DR SMART; SM00249; PHD; 1.
 DR PROSITE; PS00633; BROMODOMAIN_1; FALSE_NEG.
 DR PROSITE; PS00014; BROMODOMAIN_2; 1.
 DR PROSITE; PS50827; DDT; 1.
 DR PROSITE; PS01359; ZF_PHD_1; 1.
 DR PROSITE; PS50016; ZF_PHD_2; 1.
 KW 3D-structure; Alternative splicing; Bromodomain; Coiled coil;
 KW Nuclear protein; Transcription regulation; Williams-Beuren syndrome;
 KW Zinc-finger.
 FT DOMAIN 20 126 WAC.
 FT DDT.
 FT ZN_FING 184 1234 PHD-type.
 FT DOMAIN 1356 1426 Bromodomain.
 FT DOMAIN 306 578 Lys-rich.
 FT DOMAIN 533 586 Coiled coil (Potential).
 FT DOMAIN 768 814 Coiled coil (Potential).
 FT DOMAIN 850 893 Coiled coil (Potential).
 FT DOMAIN 1245 1283 Coiled coil (Potential).
 FT DOMAIN 1261 1273 Poly-Glu.
 FT VARSPLIC 660 663 Missing (in isoform 2).
 FT CONFLICT 14 14 K -> N (in Ref. 3).
 FT CONFLICT 22 22 L -> F (in Ref. 3).
 FT CONFLICT 136 136 K -> E (in Ref. 1).
 FT CONFLICT 191 191 N -> D (in Ref. 4).
 FT CONFLICT 298 298 Y -> V (in Ref. 4).
 FT CONFLICT 823 823 E -> R (in Ref. 3).
 FT CONFLICT 1191 1191 R -> P (in Ref. 3).
 FT CONFLICT 1354 1354 K -> M (in Ref. 2).
 FT CONFLICT 1438 1438 A -> V (in Ref. 3).
 SQ SEQUENCE 1483 AA; 170902 MW; 0CCI46FEBB954261 CRC64;
 Query Match 96.3%; Score 7674; DB 1; Length 1483;
 Beat Local Similarity 99.3%; Pred. No. 0;
 Matches 1474; Conservative 2; Mismatches 7; Indels 2; Gaps 1;
 QY 1 MAPLGRKPPFLVNPPLGCEBFFFTIPHTQEAFTREEYEALRYSEIRIWTCKTGSSQL 60
 Db 1 MAPLGRKPPFLVNPPLGCEBFFFTIPHTQEAFTREEYEALRYSEIRIWTCKTGSSQL 60
 QY 61 THKAWEEEQVAELLKEEFPANWEKLVLEWVHNTASLEKLVDTAWLEIMTKYAVGEEC 120
 Db 61 THKAWEEEQVAELLKEEFPANWEKLVLEWVHNTASLEKLVDTAWLEIMTKYAVGEEC 120
 QY 121 DFEVGKEMLVKIVKIHPLEKVDDEATEKSDGACDSPSDSKENSSQIAQHQKETVV 180
 Db 121 DFEVGKEMLVKIVKIHPLEKVDDEATEKSDGACDSPSDSKENSSQIAQHQKETVV 180
 QY 181 KEDEGRRESINDRARRSPKLPSTSLKGERKWAPPKFLPHKYDVVKLQNEDKIISNVPADS 240
 Db 181 KEDEGRRESINDRARRSPKLPSTSLKGERKWAPPKFLPHKYDVVKLQNEDKIISNVPADS 240
 QY 241 LIRTERPKNKIVRYFIHNNALRAGTGGENAPWVVEDELVKVYSLPSKFSDFLLDPKYMT 300
 Db 241 LIRTERPKNKIVRYFIHNNALRAGTGGENAPWVVEDELVKVYSLPSKFSDFLLDPKYMT 300
 QY 301 LNPSTKRKNTGSPDRKPSKSKTQNSLSPLNPKLWCHVHLKKSLSGSPKLVKNSKNSK 360
 Db 301 LNPSTKRKNTGSPDRKPSKSKTQNSLSPLNPKLWCHVHLKKSLSGSPKLVKNSKNSK 360
 QY 361 SPEEHLEEMMKOMSPNKLHTNFHIPKGGPPAKKPGKSHSDKPLKAGRSKGILNGQKSTGN 420
 Db 361 SPEEHLEEMMKOMSPNKLHTNFHIPKGGPPAKKPGKSHSDKPLKAGRSKGILNGQKSTGN 420
 QY 421 SKSPKKGKLTPTKTKMKQMTLLDMAKGTOKMTRAPNSGGTPTRTSSKPKHKHLPALHLIA 480
 Db 421 SKSPKKGKLTPTKTKMKQMTLLDMAKGTOKMTRAPNSGGTPTRTSSKPKHKHLPALHLIA 480
 QY 481 YYKENKDRKRSALSCVISKTARLLSSEDRARLPPEELRSVLQKYEYELLEHKRWASME 540
 Db 481 YYKENKDRKRSALSCVISKTARLLSSEDRARLPPEELRSVLQKYEYELLEHKRWASME 540

```
QY 541 EQRKEYLKKREELKKLKEKAKERREKEMLERLEKQRYEDQELTGKNLPAPRLVDTP 600
|
|
|
Db 541 EQRKEYLKKREELKKLKEKAKERREKEMLERLEKQRYEDQELTGKNLPAPRLVDTP 600
|
|
|
QY 601 GLPNTLPGDVAMVVEFLSCYSGLLLPDAQYPTAVSLMEALSADKGGFLYINRVLILQ 660
|
|
|
Db 601 GLPNTLPGDVAMVVEFLSCYSGLLLPDAQYPTAVSLMEALSADKGGFLYINRVLILQ 660
|
|
|
QY 661 TLLQDIAEDYGLGKMLSEIPLTLHSVSELVRLCLRRSDVQESGSDTDNKSAAFE 720
|
|
|
Db 661 TLLQDIAEDYGLGKMLSEIPLTLHSVSELVRLCLRRSDVQESGSDTDNKSAAFE 720
|
|
|
QY 721 DNEVQDEFLEKLETSFFELTSEKLIQILTALCHRLIMTYSVQDHMETRQMSAELWKR 780
|
|
|
Db 721 DNEVQDEFLEKLETSFFELTSEKLIQILTALCHRLIMTYSVQDHMETRQMSAELWKR 780
|
|
|
QY 781 LAVLKEENDKRAEKQKREMEAKNKENGKVENGLGKTDRKKIIVKFEQVDTAEADMTS 840
|
|
|
Db 781 LAVLKEENDKRAEKQKREMEAKNKENGKVENGLGKTDRKKIIVKFEQVDTAEADMTS 840
|
|
|
QY 841 AVKSRLLLAIQAKKEREIOEREMKVKLERQAEERIRKHAARAKAFQGGIAKAKLWRR 900
|
|
|
Db 841 AVKSRLLLAIQAKKEREIOEREMKVKLERQAEERIRKHAARAKAFQGGIAKAKLWRR 900
|
|
|
QY 901 TPIGTDRNHRNRYLFSDEVPGLFIEKGWVHDSIDYRFNHCKDHTVSGDEDYCPRSKKN 960
|
|
|
Db 901 TPIGTDRNHRNRYLFSDEVPGLFIEKGWVHDSIDYRFNHCKDHTVSGDEDYCPRSKKN 960
|
|
|
QY 961 LGKNASMTQHGATVAVETTPKQGNLWFLCDSQKELDELNCLHCPQIGRESOLKER 1020
|
|
|
Db 961 LGKNASMTQHGATVAVETTPKQGNLWFLCDSQKELDELNCLHCPQIGRESOLKER 1020
|
|
|
QY 1021 LEKRYODIITHSLAKPNLGLKSCDGNELNLFRLSDLIIEVATRLQKGLGVETSEF 1080
|
|
|
Db 1021 LEKRYODIITHSLAKPNLGLKSCDGNELNLFRLSDLIIEVATRLQKGLGVETSEF 1080
|
|
|
QY 1081 EARVISLEKLKDFGECVIALQASVIKKFLQGFMAPKQKRKLQSEDSAKTEEVEDEKQV 1140
|
|
|
Db 1081 EARVISLEKLKDFGECVIALQASVIKKFLQGFMAPKQKRKLQSEDSAKTEEVEDEKQV 1140
|
|
|
QY 1141 EAKVASALEKWKTAIREAQTSRMRVLLGMLDACIKWMSAENARCKVCPKKGEDDKLI 1200
|
|
|
Db 1141 EAKVASALEKWKTAIREAQTSRMRVLLGMLDACIKWMSAENARCKVCPKKGEDDKLI 1200
|
|
|
QY 1201 LCDECNKATFLCLRPALYEPDGEWQCPACQATARRNSGRNRYTEESASEDSEDS 1260
|
|
|
Db 1201 LCDECNKATFLCLRPALYEPDGEWQCPACQATARRNSGRNRYTEESASEDSEDS 1260
|
|
|
QY 1261 EEEEEEEEEEDYEVAGLRPRKTIIRGKHSVIPPAAARSGRRPGKPKHSTRRSQPKAP 1320
|
|
|
Db 1261 EEEEEEEEEEDYEVAGLRPRKTIIRGKHSVIPPAAARSGRRPGKPKHSTRRSQPKAP 1320
|
|
|
QY 1321 PVDADAEVLVLTQKSSRRQSLQKCEILHKIVKRYFSPFPFRPVTDRDAEDYDVI 1380
|
|
|
Db 1321 PVDADAEVLVLTQKSSRRQSLQKCEILHKIVKRYFSPFPFRPVTDRDAEDYDVI 1380
|
|
|
QY 1381 THPMDFQTVQNKCCSGYSRQVBFLLDMQVFTNAEYVNCRSHVLSQVKTQCLVALL 1440
|
|
|
Db 1381 THPMDFQTVQNKCCSGYSRQVBFLLDMQVFTNAEYVNCRSHVLSQVKTQCLVALL 1440
|
|
|
QY 1441 HKHLPGHVYRRKRRKKFPDLRAEDGSDSEPEAVGQSRDEDRSR 1485
|
|
|
Db 1441 HKHLPGHVYRRKRRKKFPDLRAEDGSDSEPEAVGQSR--GRQKK 1483
|
|
|
```

RESULT 2

```
BA1B MOUSE
ID BA1B MOUSE STANDARD; PRT; 1479 AA.
AC Q9227; Q9CU68;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
```

```
DE Bromodomain adjacent to zinc finger domain protein 1B (Williams-Beuren
DE syndrome chromosome region 9 protein homolog) (WBR99).
GN Name=Baz1b; Synonyms=Wbscr9;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99077764; PubMed=9858827;
RA Peoples R.J., Cisco M.J., Kaplan P., Francke U.;
RT Identification of the WBSR9 gene, encoding a novel transcriptional
RT regulator, in the Williams-Beuren syndrome deletion at 7q11.23.;
RL Cytogenet. Cell Genet. 82:238-246(1998).
RN [2]
RP SEQUENCE OF 1-482 FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Thymus;
RX MEDLINE=22354683; PubMed=12466851; DOI=10.1038/nature01266;
RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
RA Nikaado I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,
RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,
RA Balderelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
RA Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,
RA Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,
RA Datta E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,
RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,
RA Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,
RA Kanai A., Kawai H., Kawasawa Y., Kedzierski R.M., King B.L.,
RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,
RA Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,
RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,
RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,
RA Sandelin A., Schneider C., Sempile C.A., Setou M., Shimada K.,
RA Sulfana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,
RA Verrardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,
RA Wilming L.G., Wyshaw-Boris A., Yanagisawa M., Yang I., Yang L.,
RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,
RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,
RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,
RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,
RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shingawa A.,
RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
RA Birney E., Hayashizaki Y.;
RT Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.;
RL Nature 420:563-573(2002).
CC -!- FUNCTION: Forms a chromatin remodeling complex that mobilizes
CC nucleosomes and reconfigures irregular chromatin to a regular
CC nucleosomal array structure (By similarity).
CC -!- SUBUNIT: Interacts with ISWI (imitation SWI protein) to form the
CC WSTF-ISWI chromatin remodeling complex (WICH) (By similarity).
CC -!- SUBCELLULAR LOCATION: Nuclear (Potential). Accumulates in
CC pericentromeric heterochromatin during replication (By
CC similarity).
CC -!- TISSUE SPECIFICITY: Expressed in all tissues examined including
CC heart, brain, spleen, lung, skeletal muscle, kidney and testis.
CC -!- DEVELOPMENTAL STAGE: Expressed as early as day 7 and in equal
CC amounts during gestation.
CC -!- SIMILARITY: Belongs to the WAF family.
CC -!- SIMILARITY: Contains 1 bromodomain.
CC -!- SIMILARITY: Contains 1 DDT domain.
CC -!- SIMILARITY: Contains 1 PHD-type zinc finger.
CC -!- SIMILARITY: Contains 1 WAC domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC
```

DR EMBL; AF084480; AAD08676.1; -.
DR EMBL; AK017894; BAB30992.1; -.
DR PIR; T17401; T17401.
DR HSSP; Q9UIG6; I1F62.
DR MGD; MGI:1353499; Baz1b.
DR GO; GO:0005721; C:centric heterochromatin; IDA.
DR GO; GO:0000793; C:condensed chromosome; IDA.
DR GO; GO:0005634; C:nucleus; IDA.
DR GO; GO:0005515; F:protein binding; IPI.
DR GO; GO:0006333; P:chromatin assembly/disassembly; IDA.
DR GO; GO:0006338; P:chromatin remodeling; IDA.
DR InterPro; IPR001487; Bromodomain.
DR InterPro; IPR004022; DDT.
DR InterPro; IPR001965; Znf_PHD.
DR Pfam; PF00439; Bromodomain; 1.
DR Pfam; PF00628; PHD; 1.
DR PRINTS; PRO0503; BROMODOMAIN.
DR SMART; SM00297; BROMO; 1.
DR SMART; SM00571; DDT; 1.
DR SMART; SM00249; PHD; 1.
DR PROSITE; PS00633; BROMODOMAIN_1; FALSE_NEG.
DR PROSITE; PS0014; BROMODOMAIN_2; 1.
DR PROSITE; PS00827; DDT; 1.
DR PROSITE; PS01359; ZF_PHD_1; 1.
DR PROSITE; PS0016; ZF_PHD_2; 1.
KW Bromodomain; Coiled Coil; Nuclear protein; Transcription regulation;
KW Zinc-finger.
FT DOMAIN 20 126 WAC.
FT DOMAIN 605 669 DDT.
FT ZN_FING 1184 1234 PHD-type.
FT DOMAIN 1352 1422 Bromodomain.
FT DOMAIN 537 587 Coiled coil (Potential).
FT DOMAIN 769 815 Coiled coil (Potential).
FT DOMAIN 851 895 Coiled coil (Potential).
FT DOMAIN 1257 1284 Coiled coil (Potential).
SQ SEQUENCE 1479 AA; 170788 MW; DFB5816AE1C0634F CRC64;

Query Match 87.7%; Score 6989; DB 1; Length 1479;
Best Local Similarity 90.3%; Pred. No. 2.8e-289;
Matches 1433; Conservative 67; Mismatches 66; Indels 12; Gaps 7;

QY 1 MAPLLGRKFPVLNPLPGPEPFTTIPHTQEAFTREEYEARELRYSERIWTCKSGSSOL 60
Db 1 MAPLLGRKFPVLNPLPGPEPFTTIPHTQEAFTREEYEARELRYSERIWTCKSGSSOL 60

QY 61 THKEAWEEOQVAELLKEBFPNWEKLVLEMHHTASLEKLVDTAWLEIMTKYAVGEEC 120
Db 61 THKEAWEEOQVAELLKEBFPNWEKLVLEMHHTASLEKLVDSAWLEIMTKYAVGEEC 120

QY 121 DFEVGKEMLVKVIKIHPLEKVDREATEKKSDGACDSSPSKENSQIAQDHQKKEV 180
Db 121 DFEVGKEMLVKVIKIHPLEKVDREAVEKKSDGACDSSPSKENSQAQDLQKKEV 180

QY 181 KEDEGRRESINDRARRSPKLTSLKGERKWAPPKFLPHKYDVKLQNEDKIISNPADS 240
Db 181 KEDEGRRESINDRARRSPKLTSLKGERKWAPPKFLPHKYDVKLQNEDKIISNPADS 240

QY 241 LIRTERPPNKEIVRPIRINALRAGTGENAPWVSDDELVKYSLSPKSFDFLLDPKYMT 300
Db 241 LIRTERPPNKEILRPIRINALRAGTGENAPWVSDDELVKYSLSPKSFDFLLDPKYMT 300

QY 301 LNPSTKRKNTGSDPRKPSKSTDRSSLSPLNPKLWCHVHLKSLSGSPKLVKNSKNSK 360
Db 301 LNPSTKRNTGSDPRKPSKSTDRSSLSPLNPKLWCHVHLKSLSGSPKLVKNSKNSK 360

QY 361 SPEEHLEEMQWQMSPP--NKLHTNFHPIPKGPPPAKPGKHSKPLKAKGRSGKILNGKST 418
Db 361 SPEEHLEGVKIMSPNNKLH-SFHIPKGPAAKPGKHSKPLKAKGRSGKILNGKST 419

QY 419 GNSKSPKGLKTPKTKMKTLLDMAKGTQKMTAPRNSGGTPTSSPKHKLPPAALHL 478
Db 420 GNSKSPSKCVKTPKTKMKTLLDMAKGTQKMTAPRNSGGVPRSSGPKHKLPPAALHL 479

QY 479 IATYKENKDRBDRKRSALSCVISKTARLLSSBDRARLPEELRSLVQKRYELLEHKKRWASM 538
Db 480 IATYKENKDKEDKKSALSCVISKTARLLSNEDRARLPEELRALVQKRYELLEHKKRWASM 539

QY 539 SEEQKEYLKKRBELKKLKEKAKERREKEMLEKQKQYEDQELTGKNLPAFLVDT 598
Db 540 SEEQKEYLKKRQELKERLEKAKERREKEMLEKQKQYEDQELTGKNLPAFLVDT 599

QY 599 PEGLPNTLFGDVAMVVEFLSCYSGLLIPDAQYPIITAVSLMEALSADKGGFLYLNVL 658
Db 600 PEGLPNTLFGDVAMVVEFLSCYSGLLIPDAQYPIITAVSLMEALSADKGGFLYLNVL 659

QY 659 LQTLQDEIAEDYGEIGMKLSEIPLTHSVSELVRLCLRRSDVQSESGSDTDDNKDAAA 718
Db 660 LQTLQDEIAEDYGEIGMKLSEIPLTHSVSELVRLCLRRCDVQSDSGSETDDNKDSTP 719

QY 719 FEDNEVQDEFLEKLETSFEFFELTSEKLIQITALTALCHRIIMTYSVQDHMETROQMSAEIWK 778
Db 720 FEDNEVQDEFLEKLETSFEFFELTSEKLIQITALTALCHRIIMTYSVQDHMETROQMSAEIWK 779

QY 779 ERLAVLKEENDKGAEKQKREMEAKNKENGVENGLKTKDKKBIVKFEPOVDTEAEDM 838
Db 780 ERLAVLKEENDKGAEKQKREMEAKNKENGVENGLKTKDKKBIVKFEPOVDTEAEDM 839

QY 839 ISAVKSRLLAIQAKKEREIOEREMKVKLERQABEERIRKHKAAAEKAFQEGIAKALVM 898
Db 840 ISAVKSRLLSMQAKKEREIOERETKVLERAEAEERMRKHKAAAEKAFQEGIAKALVL 899

QY 899 RRTPIGTDRNHNRYWLFSDVEVPLFIEKGWVHSDIDYRFNHHCKDHTVSGDEDEYCPRSKK 958
Db 900 RRTPIGTDRNHNRYWLFSDVEVPLFIEKGWVHSDIDYRFNHHCKDHTVSGDEDEYCPRSKK 959

QY 959 ANLGKNSAMTQHGTAIVAVETTPKOGONLWFLCDKQKDELCLNCHLHPOGIRESOAK 1018
Db 960 ANLGKNSAMTQHGTAIVAVETTPKOGONLWFLCDKQKDELCLNCHLHPOGIRESOAK 1018

QY 1019 ERLEKRYQDIITHSLARKPNLGLKSCDGNQELNLFRLSDLIEVATRIKQKGLGVVEETS 1078
Db 1019 ERLEKRYQDIITHSLARKPNLGLKSCDGNQELNLFRLSDLIEVATRIKQKGLGVVEETS 1078

QY 1079 EFEARVISLEKLKDFGECVIALQASVIKKFLQGFMAPKQKRLQSEDSAKTEEVDEBK 1138
Db 1079 EFEARVISLEKLKDFGECVIALQASVIKKFLQGFMAPKQKRLQSEDSAKTEEVDEBK 1138

QY 1139 MVEEAKVASALEKWKTAIREAQTFSRMHVLTGLMDACIKWDMSAENACKVCPKKGEDDK 1198
Db 1139 MVEEAKVASALEKWKTAIREAQTFSRMHVLTGLMDACIKWDMSAENACKVCPKKGEDDK 1198

QY 1199 LILDCENKAFHLFCLRPALYEVDPGEWQCPACOPATARRNSRGRNYTEESASESDEDE 1258
Db 1199 LILDCENKAFHLFCLRPALYEVDPGEWQCPACOPATARRNSRGRNYTEESASESDEDE 1258

QY 1259 S-DEEEEBEEDYEVAGLRRLRPKRTIRGKHSVTPPAARSGRRPGKPKHSTRRQOP 1317
Db 1259 SGESEEBEEDYEVAGLRRLRPKRTIRGKHSVTPPAARSGRRPGKPKHSTRRQOP 1317

QY 1318 KAPPVDDAEVDLVLQTKSSRRSLOKCEEILHKIVKVFSPFPPEVTRDEADY 1377
Db 1318 KAPPVDDAEVDLVLQTKSSRRSLOKCEEILHKIVKVFSPFPPEVTRDEADY 1377

QY 1378 DVITHPMDFQVQNKCSGYSRVSQEFLLTMKQVFTNAEVNCRGSHVLSQVMTKEQCLV 1437
Db 1378 DVITHPMDFQVQNKCSGYSRVSQEFLLTMKQVFTNAEVNCRGSHVLSQVMTKEQCLV 1437

QY 1438 VLLHKLHFGHPYVRRKXKFPDRLAEDSGDEPEAVQSGRDEDRSRE 1485
Db 1438 VLLHKLHFGHPYVRRKXKFPDRLAEDSGDEPEAVQSGRDEDRSRE 1485

QY 1493 ALLQKHPGHPYVRRKXKFPDRLAEDSGDEPEAVQSGRDEDRSRE 1493
Db 1493 ALLQKHPGHPYVRRKXKFPDRLAEDSGDEPEAVQSGRDEDRSRE 1493

PRELIMINARY; PRT; 955 AA.
RESULT 3
Q86UJ6
ID Q86UJ6
AC Q86UJ6;

DT	01-JUN-2003	(TrEMBLrel. 24, Created)	
DT	01-JUN-2003	(TrEMBLrel. 24, Last sequence update)	
DT	01-MAR-2004	(TrEMBLrel. 26, Last annotation update)	
DE	Hypothetical protein BAZ1B (Fragment) .		
GN	Name=BAZ1B;		
OS	Homo sapiens (Human) .		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=99063792; PubMed=9847074;		
RA	Wilson R.;		
RT	"Toward a complete human genome sequence.";		
RL	Genome Res. 8:1097-1108(1998) .		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RP	Graves T., Hawkins M., Maas J.;		
RA	Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.		
RL	[3]		
RP	SEQUENCE FROM N.A.		
RA	Waterston R.;		
RL	Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.		
RN	[4]		
RP	SEQUENCE FROM N.A.		
RA	Waterston R.H.;		
RL	Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.		
RN	[5]		
RP	SEQUENCE FROM N.A.		
RA	Waterston R.;		
RL	Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.		
RN	[6]		
RP	SEQUENCE FROM N.A.		
RA	Waterston R.;		
RL	Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AC005089; AAP22332.1; -.		
DR	InterPro; IPR004022; DDT.		
DR	SMART; SM00571; DDT; 1.		
DR	PROSITE; PS00827; DDT; 1.		
KW	Hypothetical protein.		
FT	NON_TER 955		
SQ	SEQUENCE 955 AA; 110259 MW; 872A54E554424002 CRC64;		
Query Match 62.0%; Score 4940; DB 2; Length 955;			
Best Local Similarity 99.7%; Pred. No. 2.7e-202;			
Matches 952; Conservative 0; Mismatches 3; Indels 0; Gaps 0;			
QY	1	MAPLGRKPPPLVNPPLPGPEEFTTIPHTQEAFTREEYEARLERYSERIWTCKTGSSQL 60	
DB	1	MAPLGRKPPPLVNPPLPGPEEFTTIPHTQEAFTREEYEARLERYSERIWTCKTGSSQL 60	
QY	61	THKEAWEERQEAELLKEEFPAPWYKLVLEMHVHNTASLEKLVDTAWLEIMTKYAVGESC 120	
DB	61	THKEAWEERQEAELLKEEFPAPWYKLVLEMHVHNTASLEKLVDTAWLEIMTKYAVGESC 120	
QY	121	DPEVGKEMLVKYIKVIHPLEKVDDEATEKSDGACDPSDKNSSQIAQDHQKKEVTYV 180	
DB	121	DPEVGKEMLVKYIKVIHPLEKVDDEATEKSDGACDPSDKNSSQIAQDHQKKEVTYV 180	
QY	181	KEDEGRRESINDRARRSPKPLPTSLKGGKRWAPPKFLPHKYDVYKLNQEDKIISNVPADS 240	
DB	181	KEDEGRRESINDRARRSPKPLPTSLKGGKRWAPPKFLPHKYDVYKLNQEDKIISNVPADS 240	
QY	241	LIRTERPPNKEIVRYPIRHNALRAGTGENAPVVEDELVKYISLPKFSDFLLDPYKYMT 300	
DB	241	LIRTERPPNKEIVRYPIRHNALRAGTGENAPVVEDELVKYISLPKFSDFLLDPYKYMT 300	
QY	301	LNPSTKRQNTGSPDRKPSKSKTDNSLSPLNPKLWCHVHLKKSLSGSPKLVKNSKNSK 360	
DB	301	LNPSTKRQNTGSPDRKPSKSKTDNSLSPLNPKLWCHVHLKKSLSGSPKLVKNSKNSK 360	
QY	361	SPEEHLLEEMKWMSPNKLHTNFHPKGGPPAKKPGHSDKPLKAKGRSGKILNGQKSTGN 420	

Db	361	SPEEHLLEEMKWMSPNKLHTNFHPKGGPPAKKPGHSDKPLKAKGRSGKILNGQKSTGN 420	
QY	421	SKSPKKGLKTPTKMKQMTLLDMAGTQKWTAPRNSGGTPTRTSSPKHKLPPAALHLIA 480	
DB	421	SKSPKKGLKTPTKMKQMTLLDMAGTQKWTAPRNSGGTPTRTSSPKHKLPPAALHLIA 480	
QY	481	YKKNKDRDKRSALSVCVISTARTLLSSDRARLPPEELRSLVQKRYELLEHKRWASMS 540	
DB	481	YKKNKDRDKRSALSVCVISTARTLLSSDRARLPPEELRSLVQKRYELLEHKRWASMS 540	
QY	541	EORKEYLKKREELKKLEKAKERREKEMLEKQKEYEDELTKGNLPAPRLVDTP 600	
DB	541	EORKEYLKKREELKKLEKAKERREKEMLEKQKEYEDELTKGNLPAPRLVDTP 600	
QY	601	GLPNTLFGDVAMVVFSLSCYSGLLLPDAQYPTAVSLMEALSADKGGFLYLRVLVILLQ 660	
DB	601	GLPNTLFGDVAMVVFSLSCYSGLLLPDAQYPTAVSLMEALSADKGGFLYLRVLVILLQ 660	
QY	661	TLQDEIAEDYGELGKMLSEIPLTHSVSELVRLCLRRSDVQESGSDTDONKDSAAFE 720	
DB	661	TLQDEIAEDYGELGKMLSEIPLTHSVSELVRLCLRRSDVQESGSDTDONKDSAAFE 720	
QY	721	DNEVQDEFLEKLETSEFFELTSEFKLIILTALCHRLTILMTYSVQDHMETQOQSAELWKR 780	
DB	721	DNEVQDEFLEKLETSEFFELTSEFKLIILTALCHRLTILMTYSVQDHMETQOQSAELWKR 780	
QY	781	LAVLKEENDKKRAEKQKRAKEMKAKKNGKNGKGLGKTDKRIKRVKFEQVDTAEADMIS 840	
DB	781	LAVLKEENDKKRAEKQKRAKEMKAKKNGKNGKGLGKTDKRIKRVKFEQVDTAEADMIS 840	
QY	841	AVKSRRLLAIOAKEREIOEREMKVKLERQAEERIRKHAAABKAFQSGIAKAKLVMR 900	
DB	841	AVKSRRLLAIOAKEREIOEREMKVKLERQAEERIRKHAAABKAFQSGIAKAKLVMR 900	
QY	901	TPIGTDRNHNRYLWFSDEVPGLFIEKGWVHDSIDYRFNHHCKDHTVSGDEDYCP 955	
DB	901	TPIGTDRNHNRYLWFSDEVPGLFIEKGWVHDSIDYRFNHHCKDHTVSGDEDYCP 955	
RESULT 4			
Q6P1K4			
ID	Q6P1K4	PRELIMINARY; PRT; 808 AA.	
AC	Q6P1K4;		
DT	05-JUL-2004	(TrEMBLrel. 27, Created)	
DT	05-JUL-2004	(TrEMBLrel. 27, Last sequence update)	
DT	05-JUL-2004	(TrEMBLrel. 27, Last annotation update)	
DE	BAZ1B protein (Fragment).		
GN	Name=BAZ1B;		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Testis;		
RX	MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;		
RA	Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,		
RA	Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,		
RA	Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,		
RA	Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,		
RA	Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,		
RA	Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,		
RA	Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,		
RA	Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,		
RA	Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,		
RA	Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,		
RA	Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,		
RA	Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,		
RA	Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,		
RA	Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,		
RA	Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,		
RA	Krzywinski M.I., Skalska U., Smalios D.E., Schnerch A., Schein J.E.,		
RA	Jones S.J., Marra M.A.;		

RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RA Strausberg R.;
RL Submitted (JAN-2004) to the EMBL/GenBank/DBSJ databases.
DR EMBL; SC065029; AAH5029.1; -.
DR InterPro; IPR004022; DDT.
DR SMART; SM00571; DDT; 1.
DR PROSITE; PS0827; DDT; 1.
FT NON_TER 808
SQ SEQUENCE 808 AA; 93058 MW; FA06BCE0715CBF74 CRC64;

Query Match 52.2%; Score 4156; DB 2; Length 808;
Best Local Similarity 99.5%; Pred. No. 5.1e-169;
Matches 803; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 MAPILGRKPPPLVNPDPGEEPPFTIPHTQEAFTREEYEARELRYSERIWTCKSTGSSQL 60
Db |||||
Qy 1 MAPILGRKPPPLVNPDPGEEPPFTIPHTQEAFTREEYEARELRYSERIWTCKSTGSSQL 60
Db |||||

Qy 61 THKEAWEBEQEVAELLKEEFPWYKLVLEMVHNTASLEKLVDTAWLEIMTKYAVGEEC 120
Db |||||
Qy 61 THKEAWEBEQEVAELLKEEFPWYKLVLEMVHNTASLEKLVDTAWLEIMTKYAVGEEC 120
Db |||||

Qy 121 DFEVGEKMKLVKVIKIHPLKLEKDEATEKSDGACDPSKSSKENSQIAQDHQKKTWV 180
Db |||||
Qy 121 DFEVGEKMKLVKVIKIHPLKLEKDEATEKSDGACDPSKSSKENSQIAQDHQKKTWV 180
Db |||||

Qy 181 KEDEGRRESINDRARSRKLPSTLSKGERKWAPKFLPHKYDVKLQNEDKLISNVPADS 240
Db |||||
Qy 181 KEDEGRRESINDRARSRKLPSTLSKGERKWAPKFLPHKYDVKLQNEDKLISNVPADS 240
Db |||||

Qy 241 LIRTERPNNKEIVRFIRHNALRAGTGENAPWVEDELVKYSLPSKFSDFLLDPKYWT 300
Db |||||
Qy 241 LIRTERPNNKEIVRFIRHNALRAGTGENAPWVEDELVKYSLPSKFSDFLLDPKYWT 300
Db |||||

Qy 301 LNPSTKRKNTGSPDRKPKSKSTDNSSLSPLNPKLWCHVHLKSLGSPKLVKNSKNSK 360
Db |||||
Qy 301 LNPSTKRKNTGSPDRKPKSKSTDNSSLSPLNPKLWCHVHLKSLGSPKLVKNSKNSK 360
Db |||||

Qy 361 SPEEHLEEMKQMSPNKLTNPHIPKGGPPAKPKGSHDKPLKAGRSKGIILNGKSTGN 420
Db |||||
Qy 361 SPEEHLEEMKQMSPNKLTNPHIPKGGPPAKPKGSHDKPLKAGRSKGIILNGKSTGN 420
Db |||||

Qy 421 SKSPKGLKTPKTKMQLTDMAGTKQWTRAPNSGGTPTSSKPKHKLPPAALHLIA 480
Db |||||
Qy 421 SKSPKGLKTPKTKMQLTDMAGTKQWTRAPNSGGTPTSSKPKHKLPPAALHLIA 480
Db |||||

Qy 481 YYKENKREDKRSALSCVISTARTLLSSDRARLPEELRSVLQRYELLEHKRWASMS 540
Db |||||
Qy 481 YYKENKREDKRSALSCVISTARTLLSSDRARLPEELRSVLQRYELLEHKRWASMS 540
Db |||||

Qy 541 EQRKEYLKKRBEELKKLEKAKERREKEMLEKQRYEDELTKGNLPAFLVDTPPE 600
Db |||||
Qy 541 EQRKEYLKKRBEELKKLEKAKERREKEMLEKQRYEDELTKGNLPAFLVDTPPE 600
Db |||||

Qy 601 GLPNTLFGDVAMVVEFLSCYGLLLPDAQYPTTAVSLMEALSADKGGFLYLNRLVILLO 660
Db |||||
Qy 601 GLPNTLFGDVAMVVEFLSCYGLLLPDAQYPTTAVSLMEALSADKGGFLYLNRLVILLO 660
Db |||||

Qy 661 TLLQDEIAEDYGEKMGKLSIEPLTLHVSSELVRLCLRRSDVQEESEGSDDTDNDKSAAFE 720
Db |||||
Qy 661 TLLQDEIAEDYGEKMGKLSIEPLTLHVSSELVRLCLRRSDVQEESEGSDDTDNDKSAAFE 720
Db |||||

Qy 721 DNEVQDEFLEKLETSFFELTSEEKLIITALCHRLMTYSVQDHMETQOQMSBLWKER 780
Db |||||
Qy 721 DNEVQDEFLEKLETSFFELTSEEKLIITALCHRLMTYSVQDHMETQOQMSBLWKER 780
Db |||||

Qy 781 LAVLKEENDKKRAEKQKRKEMEAKNKE 807
Db |||||

Db 781 LAVLKEENDKKRAEKQKRKEMEAKNKE 807

RESULT 5
Q8UVR4
ID Q8UVR4 PRELIMINARY; PRT; 1079 AA.
AC Q8UVR4;
DT 01-MAR-2002 (TrenBLrel. 20, Created)
DT 01-MAR-2002 (TrenBLrel. 20, Last sequence update)
DE 01-MAR-2004 (TrenBLrel. 26, Last annotation update)
OS Williams syndrome transcription factor (fragment).
OC Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21668941; PubMed=11809820;
RA MacCallum D E., Losada A., Kobayashi R., Hirano T.;
RT "ISWI remodeling complexes in Xenopus egg extracts: identification as
RL major chromosomal components that are regulated by INCENP-aurora B.";
Mol. Biol. Cell 13:25-39(2002).
DR EMBL; AF412333; AAL60161.1; -.
DR HSSP; Q9UIG0; 1F62.
DR GO; GO:0000151; C:ubiquitin ligase complex; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0004842; F:ubiquitin-protein ligase activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0016567; P:protein ubiquitination; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR Pfam; PF00439; Bromodomain; 1.
DR Pfam; PF00628; PHD; 1.
DR PRINTS; PR00503; BROMODOMAIN.
DR SMART; SM00297; BROMO; 1.
DR SMART; SM00184; RING; 1.
DR SMART; SM00249; PHD; 1.
DR PROSITE; PS50014; BROMODOMAIN_2; 1.
DR PROSITE; PS50827; DDT; 1.
DR PROSITE; PS01359; ZF_PHD_1; 1.
DR PROSITE; PS50016; ZF_PHD_2; 1.
DR PROSITE; PS50089; ZF_RING_2; 1.
FT NON_TER 1
SQ SEQUENCE 1079 AA; 124503 MW; 2AF4689FFFE9E83DF CRC64;

Query Match 46.9%; Score 3740; DB 2; Length 1079;
Best Local Similarity 66.1%; Pred. No. 3.5e-151;
Matches 737; Conservative 140; Mismatches 186; Indels 52; Gaps 14;

Qy 383 HIPKGPAPKPKGSHDKPLKAGRSKGIILNGKSTGNSKSPKGLKTPKTKMQLTLD 442
Db |||||
Qy 1 NLNKKG----SUTGKSKKKKPKNGKSKQ-VLNGQKVSQKTRSPKAVKGP--KLKQNTLLD 54
Db |||||

Qy 443 MAKGTQKMTAPRNSGGTPTSSKPKHKLPPAALHLIAYKENKREDKRSALSCVIST 502
Db |||||
Qy 55 MAKSTPKVSRQKSGSSTPRSSSKPKHKLPPAALHLISYPRDNKNREDKRSALSISKV 114
Db |||||
Qy 503 ARLSSSDRARLPEELRSVLQRYELLEHKRWASMSSEQRKEYLKKRBEELKKLEKKA 562
Db |||||
Qy 115 ARMLTAEDRKRLPDDLQELVQKRYDLLEHRRKQWAVMTBEQREYVNRKREALKVRICKA 174
Db |||||
Qy 563 KERRKEMLEKLEKQRYEDELTKGNLPAFLVDTPPEGLNNTLFGDVAMVVEFLSCYSG 622
Db |||||
Qy 175 RERKQKEERLEKQRYEDELTKGLDPTFKLVDTPEGLPNALFGDVAMVVEFLSGYSD 234
Db |||||
Qy 623 LLLPDAQYPTTAVSLMEALSADKGGFLYLNRLVILLOTLQDEIAEDYGEKMGKLSIEP 682
Db |||||
Qy 235 LLLPDAQYPTTAVSLMEALSADKGGFLYLNRLVILLOTLQDEIAEDYGEKMGKLSIEP 294
Db |||||
Qy 683 LTLHVSSELVRLCLRRSDV---QEESEGSDDTDNDKSAAFEDNEVQDEFLEKLETSFF 739
Db |||||
Qy 295 LVLSASSELVRLCLRRKSDSPAGESDSTKGDSDSVGSVAVYQDDEVEDEYLEKLETSFFE 354
Db |||||

```
QY 740 LTSEEKLOILTALCHRLTMTYSVODHMETROQMSAELWKERLAVLKEENDKKBAEKQK 799
Db 355 LTTEEKLLHLLAALCHRLTMTYSVQDHVDAKQQRSGELWKERLALKEENDKKBAEKQK 414
QY 800 EM-EAKNKENGKVGGLGKTDRKRIIVKPPQVDTEADMSIAVSKRRLLAIQAKEREI 858
Db 415 EOGTVRPKE-----EVQAAKIVKQE--KINTQDNDAEDMSIAVSKRRLLQMAQKEKE 468
QY 859 QEREMKVLRQAEEERIRKHAAAEKAPQEGIAKAKLVNRRPTIGTDRNHNRYWLFSD 918
Db 469 HEXLTKERIETRETEERSRKAKAEKAFHEGIAKAKLVLRSPFLGTDNRNHNRYWLFSD 528
QY 919 VPGLFTEKGWVHDSIDYRPNHCKDHTVSGDEY-----CPRSKANLGNKVASWNTQ 970
Db 529 VPGLYIEKGWVHDSINRYSPESQKSEQDAEESDANSIGCPDDSTQREEKHA----- 583
QY 971 HGTATEVAVETTPKQGNLWFLCDSQKELDLNCLHPQIGRESOLKERLEKRYODI1H 1030
Db 584 -----ETTVPKQGNLWFLCDTQKELDLNCLHPQGFRESQKLERLQRYODI1H 634
QY 1031 SIHLARKPNLGLKSCDGNQELLNFLRSDLIIEVATRLQKGLGYVEETSFEARVISLEKL 1090
Db 635 SIHLARKQNLGLKTCGQQLNFLRSDIIEVATRLQKGLGYLDDTTEFEAKVTFENL 694
QY 1091 KDFGECVIALQASVIKFLQGFMAPKQKRLKLOSE-SAKTERVDEKKNVEAKVASAL 1149
Db 695 KDFGECIVLQAAVIKFLQGFMAPKQKRLKQSEEAQAEQDSEKQMAEAKVASAV 754
QY 1150 EKKWTAIRAEQTSRHHVLLGMDACIKWDMSENARCKVCPKKGDDDLILCDECNKAP 1209
Db 755 EKKWVAIRDAQTSRHHVLLGMDACIKWDMSENARCKVCPKKGDDDLILCDECNKAP 814
QY 1210 HLFCLRPALYEVDPGEMQCPACQATARRNSRGRNTYTESASDSDDSDDESEEEEEE 1269
Db 815 HLFCLRPVLFNI PDGEMWLPACQATARRNSRGRNYAEDSTQDDESEEESEEEEEE 874
QY 1270 --BEEDYEVAGLRLPRKTIROKHSVIPPAARSGRRPKKPHSTRSQPKAPVDDAEV 1327
Db 875 SDBEESEQEMGQRLRSRAAKKPG---RPTRRGRPPKNNTHS-RVSRQRYVEDTEADV 930
QY 1328 DELVLOTKSSRSQSLQCEILHKIKVYRFSWPFPVTFDEADYDVVITHPMDQ 1387
Db 931 EEMVRQKPTSRKQNGEFOKCEILAKIKYRFSWPFPFNDAEIDYTKVVTTPMDQ 990
QY 1388 TVQNKSCGSYSRVQBFELDMKVFTNAEYVNCGRSHVLSVMKVCQLVLIHLKHPGH 1447
Db 991 TWQSKSCGSYQTVQFNLNKLKLVFCNTELYEAGSSQLSCLTEQCCARDLLGKLP 1050
QY 1448 PYVRKRPKPPDLAEDEGDSPEAVQSGRDEDR 1482
Db 1051 TYQRRHRK-----HQSPEPEPETANPGRGRKQK 1078

RESULT 6
Q8CAU9 PRELIMINARY; PRT; 657 AA.
AC Q8CAU9;
DT 01-MAR-2003 (TEMBLrel. 23, Created)
DT 01-MAR-2003 (TEMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TEMBLrel. 26, Last annotation update)
DE Mus musculus 16 days neonate thymus cDNA, RIKEN full-length enriched
DE library, clone:A130044E01 product:bromodomain adjacent to zinc finger
DE domain, 1B, full insert sequence. (Fragment).
GN Name=Baz1b;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
[1]
RN SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Thymus;
RC MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA Carninci P., Hayashizaki Y.;
```

```
RT "High-efficiency full-length cDNA cloning.";
RL Meth. Enzymol. 303:19-44(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Thymus;
RC MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RX RIKEN FANTOM Consortium;
RA "Functional annotation of a full-length mouse cDNA collection.";
RT Nature 409:685-690(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Thymus;
RC The FANTOM Consortium;
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Thymus;
RC MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RX Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.,
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Hazada A.,
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
RT sequencing pipeline with 384 multicapillary sequencer.";
RL Genome Res. 10:1757-1771(2000).
RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Thymus;
RC MEDLINE=2030913; PubMed=11076861; DOI=10.1101/gr.152600;
RX Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RA Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M.,
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Hazada A.,
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
RT sequencing pipeline with 384 multicapillary sequencer.";
RL Genome Res. 10:1757-1771(2000).
RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Thymus;
RC Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
RA Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,
RA Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasekawa T.,
RA Katoh H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,
RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
RA Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,
RA Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
RA Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,
RA Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,
RA Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;
RL Submitted (JUL-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL; AK037737; BAC29862.1; -.
DR HSSP; O9UIG0; 1F62.
DR MGD; MGI:1353499; Baz1b.
DR CO; GO:0005721; Centric heterochromatin; IDA.
DR GO; GO:0000793; Condensed chromosome; IDA.
DR GO; GO:0005634; C:nucleus; IDA.
DR GO; GO:0005515; F:protein binding; IPI.
DR GO; GO:0006333; P:chromatin assembly or disassembly; IDA.
DR GO; GO:0006338; P:chromatin remodeling; IDA.
DR InterPro; IPR001487; Bromodomain.
DR InterPro; IPR011011; FYVE PHD Znf.
DR InterPro; IPR001965; Znf PHD.
DR InterPro; IPR001841; Znf ring.
DR Pfam; PF00439; Bromodomain; 1.
DR Pfam; PF00628; PHD; 1.
DR PRINTS; PR00503; BROMODOMAIN.
DR SMART; SM00297; BROM; 1.
DR SMART; SM00249; PHD; 1.
```

```
DR SMART: SM00184; RING; 1.
DR PROSITE; PS50014; BROMODOMAIN_2; 1.
DR PROSITE; PS01359; ZF_PHD_1; 1.
DR PROSITE; PS50016; ZF_PHD_2; 1.
DR PROSITE; PS50089; ZF_RING_2; 1.
FT NON_TER 1
SQ SEQUENCE 657 AA; 76089 MW; A87C106F4B8BA480 CRC64;

Query Match 38.0%; Score 3029.5; DB 2; Length 657;
Best Local Similarity 87.4%; Pred. No. 3.5e-121;
Matches 581; Conservative 36; Mismatches 39; Indels 9; Gaps 5;

QY 822 KRIVFEPQVDTAEADMSIAVKSRLLLAIQAQKEREIOEREMKVKVLEQAEEERIKKKA 881
Db 1 KEIVKIQEVEADMSIAVKSRLLSQAQKREIOERETKYVLEEAEEERKMKKA 60
QY 882 AAEKAFQEGIAKAKLVMBRTPTGTDNRNHNRYLWFSDEVPGLFIEKGWVHDSIDYRFNHC 941
Db 61 AAEKAFQEGIAKAKLVMBRTPTGTDNRNHNRYLWFSNEVPLFIEKGWVHNSIDYRFKHR 120
QY 942 KHTVSGDEDYCPRSKANLGNASWNTQHGTAIVETTPKQGNLWFLCDSQKELD 1001
Db 121 KDSNLPLDDYCPRSKANLGNASVNAHGGPALE-AVETTVPKQGNLWFLCDSQKELD 179
QY 1002 ELLNCLHPQIGRESOLKERLEKRYODIITHSLARKPNLGLKSCDGNQELNFLRSLIE 1061
Db 180 ELLSCLHPQIGRESOLKERLEKRYQETHTSIYLARKPNLGLKSCDGNQELNFLRSLIE 239
QY 1062 VATRLQKGLGVETSTSEFARVISLEKLDKFGECVIALQASVINKFLQGFMAPKQKRR 1121
Db 240 VATRLQKGLGVWEGTSEFARVISLEKLDKFGECVIALQASVINKFLQGFMAPKQKRR 299
QY 1122 LOSDSAKTEEVDEBKWVEAKVASALEKWKTAIREAQTFSRMHVLLGMLDACIKWMS 1181
Db 300 LOSDSTKSEEVDEBKWVEAKVASALEKWKTAIREAQTFSRMHVLLGMLDACIKWMS 359
QY 1182 AENACKVCPKKGDDKLILDCENKAFHLCRLPALYVDPDGEWQCPACAPATARRNSR 1241
Db 360 AENACKVCRKKGDDKLILDCENKAFHLCRLPALYVDPDGEWQCPACOPPTARRNSR 419
QY 1242 GRNYTEESASESDEDES-DEEEEEEEDYEVAGLRRLPRKTIIRGKHSHVPPAAR 1300
Db 420 GRNYTEESTSESGDESGEEEEEEDYEVAGLRRLPRKTIIRGKHSHVPPAAR 478
QY 1301 SORRGCKPHSTRSQPKAPPVDDDAEVLVLTQTRSSRRSLELOKCEEILHKLVKVRP 1360
Db 479 PGRPFCKSHPARSRPK-----DDPEVDVLVLTQTRISRRSLELOKCEDILHKLVKVRP 534
QY 1361 SWPFREPVRDEADYDVITHPMDPQTVQNKSCGSYSRSVQEFITDMKQVFTNAEVVNC 1420
Db 535 SWPFREPVRDEADYDVIEHPMDPQTVQNKSCGNTSRVQEFITDMKQVFNANAEVNC 594
QY 1421 RGSVLSVCWVTEQCLVLLKHLPGHPVVRKRKKFPDRLAEDSGDSEPAVGQSRDED 1480
Db 595 RGSVLSVCWKEQECLLALLOKHLPGHPVVRKRKKFPDRLAEDSGDSESVGQSR--G 652
QY 1481 RRSRE 1485
Db 653 RRQKK 657

RESULT 7
Q6GQ06 PRELIMINARY; PRT; 777 AA.
AC Q6GQ06;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE LOC443594 protein (fragment).
GN Name=LOC443594;
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae;
```

```
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RL and mouse cDNA sequences.";
RN Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RP [2]
RC SEQUENCE FROM N.A.
RX TISSUE=Embryo;
RA MEDLINE=22341132; PubMed=12454917; DOI=10.1002/dvdy.10174;
RA Klein S.L., Strausberg R.L., Wagner L., Pontius J., Clifton S.W.,
RA Richardson P.;
RT "Genetic and genomic tools for Xenopus research: The NIH Xenopus
RL initiative.";
RN Dev. Dyn. 225:384-391 (2002).
RP [3]
RC SEQUENCE FROM N.A.
RX TISSUE=Embryo;
RA Klein S., Gerhard D.S.;
RL Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC072944; AAH72944.1; -.
DR InterPro; IPR004022; DDT.
DR SMART; SM00571; DDT; 1.
DR PROSITE; PS50827; DDT; 1.
FT NON_TER 777
SQ SEQUENCE 777 AA; 89051 MW; 8DBABF7365DE2087 CRC64;

Query Match 36.1%; Score 2878.5; DB 2; Length 777;
Best Local Similarity 69.5%; Pred. No. 1.1e-114;
Matches 567; Conservative 103; Mismatches 91; Indels 55; Gaps 13;

QY 1 MAPLLGRKPPPLVNPPL-----PGEPPFTTPTHTQEAFTRTREEYEARLERYSERIWTCKS 54
Db 1 MAPLLGRPPPLVNPPLSEAAATGEGEEVYMIHSEKAFPSREYESRLRYAERIWTCKS 60
QY 55 TGSSQLTHKEAWEEEOEVAELIKKEFPANWYKLVLEWHHTATSLKLVDTAWLEIMTKY 114
Db 61 TGSSQLTHKEAWDEEQEVAELLKEEFPVYKQVLEWHHTNTISLDKLVDSQSWMEIMTKY 120
QY 115 AVGECEDFEVGKEMKLVKVIKHPLEKVDDEATEKKSQDGCADSPSSDKENSSQIAQDHQ 174
Db 121 ADGECEDFEVGPEKYLRAKIVKVPLEK-EQEAQSEKKEGSCDPSDDKSNKVAQDIQ 179
QY 175 KKETVVKDEGRRESI-----NDRARRSPKRLPTSLKKGERKWAAPPKPLPHKYDVKLONE 229
Db 180 -----LKEESNRLESLSLRESDRARRSPKRLPTSLKKEKKWVPPKPLPHKYDVKLONE 234
QY 230 DKILSNVPADSLIRTERPPNKEIVRYFIRHNALRAGTGENAPWVVEDELVKKYSLPSKFS 289
Db 235 DKVISFVPEVDSLYRSEPPNKEILRYFIRHNALRIGTGENAPWVVEDELVKKYSLPSKFS 294
QY 290 DFLDPPKYMTLNPS--TKRKNTGSPORKPSKSKTQNSLSLSSPLNPKLWCHVHLKKSL 347
```

Db 295 DFLDDPHKWTNLPNSATKRSKSGSDQKPAKSK-----KSPLSFSSWSLANLKKT-- 346
Qy 348 GSPKLVKNSKNSKSPBEHLEEMKMSPNKLHFNPHPKKGPAPKPKGHKSHDKPLAKGR 407
Db 347 -----AVNSSE---EEMQLMIGAN-----LNKKG-----SIGKSKDKKPKNGK 384
Qy 408 SKGILNGKSTGNSKSPKGLKTPKTKMOMTLLDMAKGTQKQVTRAPRNSGGTPTRTSSKP 467
Db 385 SQ-VLNGOKISAKTRSPKGLKSP--KLQMTLLDMAKSTPKVSRQAQKGSNTPRSSKP 441
Qy 468 HKHLPALHLIAIYKENDREDKRSALSCVISTARLLSLSSDRRLPEELRSLVOKRYE 527
Db 442 NKTLPPAALHLIYKDNKNREDKRSALSLISKVARMJSAEDKKLPDDLQELVOKRYE 501
Qy 528 LLEHKRWASMSSEQRKEYLKKREELKKLKEKAKERKEKEMLERLEKQRYEDQELTG 587
Db 502 LLEHRQWAVMTEQREYMRKREALKARIKEKTRERKQKREERLEKQRYEDQDLTG 561
Qy 588 KNLPAFLVDTPLGLPNTLFGDVAMVVEFLSCYSGLLLPDPAQPIPTAVSLMEALSADKGG 647
Db 562 KSLPTEKLVDTPLGLPNALFGDVAMVIEFLSGYSDLLLPDQGPVPTAVSLMEALAAKGG 621
Qy 648 ELYLNEVLVLLQTLQDDEADYDGLGKWLSEIPLTLHSVLSVLPCLRRSDV---QBE 704
Db 622 FMYLNRGLVLLQTLQDDEADYDGLGKWLSEIPLTLHSASVLSVLPCLRRKSDSPAGENE 681
Qy 705 SEGSDTDNNKSAAFEDNEVDQEFLEKLETSEFFELTSEKLOILTALCHRLIMTYSVOD 764
Db 682 STEKGDSEGSNAVQDDEVDYLEKLETSEFFELTSEKLOILTALCHRLIMTYSVOD 741
Qy 765 HMETRQMSAELWKLAVLKEENDKRAEKQKKE 800
Db 742 HVDKQKQSGELWKLAVLKEENDKRAEKQKKE 777

RESULT 8
BAIA_HUMAN STANDARD; PRT; 1556 AA.
ID BAIA_HUMAN STANDARD; PRT; 1556 AA.
AC Q9NRL2; Q9N215; Q9P065; Q9UIG1; Q9Y3V3;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Bromodomain adjacent to zinc finger domain protein 1A (ATP-utilizing
DE chromatin assembly and remodeling factor 1) (hACF1) (ATP-dependent
DE chromatin remodelling protein) (Williams syndrome transcription
DE factor-related chromatin remodeling factor 180) (WCRF180) (hWALpi)
DE (CHRAC subunit ACF1) (HSPC317).
GN Name=BAZ1A; Synonyms=ACF1, WCRF180;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1), AND MASS SPECTROMETRY.
RC TISSUE=Cervical carcinoma;
RX MEDLINE=20341093; PubMed=10880450; DOI=10.1093/emboj/19.13.3377;
RA Poot R.A., Dellaire G., Huelsmann B.B., Grimaldi M.A., Corona D.F.V.,
RA Becker P.B., Bickmore W.A., Varga-Weisz P.D.;
RT "HuCHRAC, a human ISWI chromatin remodelling complex contains hACF1
RT and two novel histone-fold proteins.";
RL EMBO J. 19:3377-3387(2000).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RX MEDLINE=20122567; PubMed=10655480; DOI=10.1073/pnas.97.3.1038;
RA Bochar D.A., Savard J., Wang W., Lafleur D.W., Moore P., Cote J.,
RA Shiekhattar R.;
RT "A family of chromatin remodeling factors related to Williams syndrome
RT transcription factor";
RL Proc. Natl. Acad. Sci. U.S.A. 97:1038-1043(2000).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Testis;
RX MEDLINE=20130112; PubMed=10662543; DOI=10.1006/geno.1999.6071;

RA Jones M.H., Hamana N., Nezu J.-I., Shimane M.;
RT "A novel family of bromodomain genes.";
RN Genomics 63:40-45(2000).
RN [4]
RP SEQUENCE OF 1-541 FROM N.A. (ISOFORM 1).
RC TISSUE=Blood;
RA Ye M., Zhang Q.H., Zhou J., Shen Y., Wu X.Y., Guan Z.Q., Wang L.,
RA Fan H.Y., Mao Y.F., Dai M., Huang Q.H., Chen S.J., Chen Z.;
RT "Human partial CDS from cd34+ stem cells.";
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE OF 746-1556 FROM N.A. (ISOFORM 1).
RC TISSUE=Uterus;
RA Wambutt R., Heubner D., Mewes H.-W., Gassenhuber J., Wiemann S.;
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: May play a role in transcriptional regulation. May be
CC involved in the formation or maintenance of heterochromatin
CC playing a critical role in developmental control.
CC -!- SUBUNIT: Together with CHRAC1, POE3 and ISWI/SNF2H proteins, it
CC forms the ISWI chromatin remodelling complex, CHRAC.
CC -!- SUBCELLULAR LOCATION: Nuclear. May target the CHRAC complex to
CC heterochromatin.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=1;
CC IsoID=Q9NRL2-1; Sequence=Displayed;
CC Name=2;
CC IsoID=Q9NRL2-2; Sequence=VSP_000551;
CC -!- TISSUE SPECIFICITY: Highly expressed in testis and at low or
CC undetectable levels in other tissues analyzed.
CC -!- MISCELLANEOUS: Stimulated by double-stranded DNA and nucleosomal
CC DNA.
CC -!- SIMILARITY: Belongs to the WAL family.
CC -!- SIMILARITY: Contains 1 bromodomain.
CC -!- SIMILARITY: Contains 1 DDT domain.
CC -!- SIMILARITY: Contains 1 PHD-type zinc finger.
CC -!- SIMILARITY: Contains 1 WAC domain.
CC -!- CAUTION: Ref.4 sequence differs from that shown due to multiple
CC frameshifts starting at position 382.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF213467; AAF70601.1; -;
DR EMBL; AF221130; AAF32366.1; -;
DR EMBL; AB032252; BAA89209.1; -;
DR EMBL; AF161435; AAF28995.1; ALT_SEQ.
DR EMBL; AL050089; CAB43261.1; -;
DR PIR; T08738; T08738.
DR HSSP; Q9UIG0; 1F62.
DR Genew; HGNC:960; BAZ1A.
DR MIM; 605680; -;
DR InterPro; IPR001487; Bromodomain.
DR InterPro; IPR004022; DDT.
DR InterPro; IPR001965; Znf_PHD.
DR InterPro; IPR001841; Znf_PHD.
DR Pfam; PF00439; Bromodomain; 1.
DR Pfam; PF02791; DDT; 1.
DR Pfam; PF00628; PHD; 1.
DR PRINTS; PR00503; BROMODOMAIN.
DR SMART; SM00297; BROMO; 1.
DR SMART; SM00571; DDT; 1.
DR SMART; SM00249; PHD; 1.
DR SMART; SM00184; RING; 1.
DR PROSITE; PS00633; BROMODOMAIN_1; 1.
DR PROSITE; PS50014; BROMODOMAIN_2; 1.
DR PROSITE; PS50827; DDT; 1.
DR PROSITE; PS01359; ZF_PHD_1; 1.

Query Match	8.3%;	Score 662;	DB 2;	Length 1476;	
Best Local Similarity	18.3%;	Pred. No. 4.8e-20;			
Matches 329;	Conservative 256;	Mismatches 502;	Indels 714;	Gaps 56;	
QY	3	PILGRKPPVLPNPLGEEEPF-----FTTPHTQEAFTREEYEARLERYSERIWTCKSTG	56		
Db	2	PICKREGFDLNQKQGNFTDNDQVFCYIITKRIFRDYEHFRHVIVNSTVMOCEATG	61		
QY	57	SSQLTHKEAWEEOEVAELLKEEFPAPWYEKLIVLEMHNTASLEKLVDTAWLEIWK- YA	115		
Db	62	KENLYIEEAVKSER-AARKMEOFKQSLRSPVLLVVEHAQQSAVNTLNVIAKFLKRYF	120		
QY	116	VGEEDCFEYKGBKMLKVIVKIHLEKVDDEATEKSGDGDSPSDKENSQIAQDHQK	175		
Db	121	IGEEVSQAQKNATYVTLGVKL-----DKMPEPLNGIYED	156		
QY	176	KETVVKEDGRRRESINDRARRSPKLTSLKGERKWAAPPKPLPHKYDVKLQNEDKIISN	235		
Db	157	TDNLVY-----RURPNKGDSAE-----LD	176		
QY	236	VPADSLIRTERPNKEIVRYFIRHNALRAGTGENAPWVVEDELVKYSLPSKFSDFLLDP	295		
Db	177	LPFRQLRRRMEFLENLSNFIKSNVSRV-----DGLLR--PKPEAYKQVYVDP	223		
QY	296	YKYMTLNPSTKRKNTGSPDRPKSKSDTNSLSPLNPKLWCHVHLKSLSGSPKVKVN	355		
Db	224	-----GWNFST-----IFIGMPRYSAPAKI- 244			
QY	356	SKNSKSPBEHLEMMKMSPNKLTNTHFI PKGPPAKPGKHSKDPKAKGRSKGILNQ	415		
Db	245	-----KPDGK-----KOSTLNKVIAGE	262		
QY	416	KSTGNSKSPKGLKTPKTQWKQMTLLDMAKGTQKQWTRAPRNSGGTPRTSSKPKHLPPAA	475		
Db	263	ATAAKSKAKK-----	273		
QY	476	LHLIAYKENKDREDKRSALSCVISKTARLLSSEDRARLPBELSLVQKRYELLEHKKRW	535		
Db	274	-----SDAKSLAEELRV-----	286		
QY	536	ASMSBEORKEYLKKREELKKLKEKAKERRERKEMLERLEKQ-----KRYEDQELTG-KN	589		
Db	287	-----KREKEAKLIELEKQKAEKQALIERVENECNLLQKTDLERTDQKV	333		
QY	590	LPAPRLVDTPEGLPNTLFGDVAMVVEFLSCYSGLL--LPDAQYPTAVSLMEALSA-DKG	646		
Db	334	LPYRQIVTL--LPEHLGDAFMREFMHTYTGLLSGIEVFRQNLSEFYEMTRALTAREIA	391		
QY	647	GLYLNRLVILLQTL-----ODSIAEDY-----GELG	675		
Db	392	G--PLSDILLVLGTVFDLQKEEBEBCAVTYLDRAAQTOEPYWSMAQAASHLYAKRHFS	449		
QY	676	MKLSEIPLTLHVSSELVRLCLRRSD--VOESEG-----SDTDDNKSAAAFEDNEVDQEF	728		
Db	450	FKVNEPLDALTLNEVRLHLLGSAFVNEKAERWVRVYRNGYSKEDPDGLELRLEHSHI	509		
QY	729	LEKLETSEFFELTSEKLIQLTALCHRIILMTYS-VQDHMETROQMSAELWKERLAVLKEE	787		
Db	510	LRILKNHSYQKFDKIMILLIRCLMSQI-MTYSGTINLIEERMEQOTAKARQDLRALVGE	568		
QY	788	NDKRAEKQKREMEAKNKENGKGLGKTDKRRKRVKFEVQVDTAEADM- ISAVKSR	846		
Db	569	NKRLAAVEINRKL-----TQMHLLEVNGVEPEK	597		
QY	847	LIAIAQKRETOEREMKVLQERQAEERIRKHAAAEKAFQEGIAKAKLVNRRPTIGTD	906		
Db	598	REALVEKLKKSIAE-----LHAQSDQOH-RXHELQMLK-----LHSLFNFLVYLGMD	644		
QY	907	RHNHNYWLFSDVPLGFIKGVHVSIDYRFNHHCKDHTVS-----	947		
Db	645	RCRYKYVYL-ESMPGIFVEHS--PDSLD-----TCLQPIITNKSQIETRQOSALPKNRKD	696		

QY	948	-----GDEDYCPRSKANKLGNASMNTOH--GTATEVAVETTT	984		
Db	697	LRVILLKLYGDDEK-KTKKKAKHSLKENQEHRLNGSAEPMDSPEAPTHFELLMC	755		
QY	985	-----KQOQ-NLWFLCDSQKELDELLNCLHPOGIBESQ	1017		
Db	756	SGDKRSCIVHDSRNGRQORWAYIYKABEIDELIKALNPNGRLREYELLQBELSVLRSIIOH	815		
QY	1018	-----KERLEKRYQDIHHSIHLARK--PNLGLKSCDGNQOELLNF	1054		
Db	816	AKTCPVLLLSLENETMRKFKFMAAMES-ETNRKYGEANFGLPNGTDLNVEWRHLVDRIIQ	874		
QY	1055	-----LRSDLI-----EVATRLQKGLGYVE	1079		
Db	875	FENDIYTGDLGRLLKVKDMEKWRSLLGNGNYDAQCKLQWGGGKLEBAGSDNESHETHEE	934		
QY	1080	-----FEARVISLEKLKO	1100		
Db	935	DGALLGKYARKPYRDPGMYLAASADTKPLPDDDEDOHTNAVSPIAVHNNASALLQV	994		
QY	1101	OASVIKKFLQ--GFMA--PKOKRRKLQESDSAKTEEVDEEKKWVEEAKVASALEKWKTA	1155		
Db	995	EQAIKGRFLKEPYGMKKWDPKQEAALKACD-----SRLHQWEVS	1033		
QY	1156	IREAQTFSRMHYLLGLMDACIKWMSAENARCKVCPKGEDDKLILCDECNKAFHFLCLR	1215		
Db	1034	LMESTSAQVFLHLNLHDCIQWRRSTNKSCLKVCRRGSDPEKMLLDCDECNAGTHMFLCK	1093		
QY	1216	PALYVDPGEWQCPACQATARRNSGRNYTESASE-----DSEDDSEDEEEEEEE	1269		
Db	1094	PKLRSVPPGNWYCNDCVKSGLSNGQNEKDKQATKKKGFIVEEDEDEATDEEBEKKD	1153		
QY	1270	EEEDYEVAGLRRLPRKTIRGKHSV--IPPAARG-----RRPGKKPHSTRRSQPK	1318		
Db	1154	DMTDEDAEHEKDEHDEVEDDESVTSTPSSSRVNGRIILRRPTRTSRLTSKEIEEHA	1213		
QY	1319	-----APPVDDAEVDLVLQT-----	1334		
Db	1214	QEDVDSGVDDASLTAGEDTTEDSDEBKVCQKCFYDGGEIKVCQCLFFHLECVHLKR	1273		
QY	1335	-----	1334		
Db	1274	PPRTDFVCKTCKMPQPRRRHSNNMGDHRDEEBEPKAKRPNRSURLCIDKTARPSNGNN	1333		
QY	1335	-----KRSRRSQS-----LELQKCEIILHKIVKYPSPWFRPVPTRDEADY	1376		
Db	1334	NNNNNNSSVNNHRRSGRTNEHPLNSAALYDLLEQIMKKAAWFLRPLVLTSEVPDY	1393		
QY	1377	YDVITHPMDFQTVQNKSCGYSYSVQBEFLTDMKQVFTNAEVYNCRGSHVLSMVKTEQCL	1436		
Db	1394	HQIITPMDLAKIKSLANWGAQLNEELLSDIQLVFRNCNCLYVNEGNIYDAGCQLERFV	1453		
QY	1437	V 1437			
Db	1454	I 1454			
RESULT 11					
Q9YOW1 PRELIMINARY; PRT; 1476 AA.					
AC	Q9YOW1				
DT	01-NOV-1999 (Tremblrel. 12, Created)				
DT	01-NOV-1999 (Tremblrel. 12, Last sequence update)				
DT	01-MAR-2004 (Tremblrel. 26, Last annotation update)				
DE	ATP-dependent chromatin assembly factor large subunit.				
GN	Name=Actf1;				
OS	Drosophila melanogaster (Fruit fly).				
OC	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;				
OC	Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;				
OC	Ephydroidea; Drosophilidae; Drosophila.				
OX	NCBI_TaxId=7227;				
RN	[1]				
RP	SEQUENCE FROM N.A.				

QY 1335 -----KSSRRQS-----LELOKCEIILKIVKVFSPFPVTR 1370
Db 1328 PGNNGNNNNNNSSVNNHRRSGRTNEHMLNSAALYDLLEQIMKHAAMPFLRPVLT 1387
QY 1371 DEAEYDYVITHPMDFQVONKSCGYSVDFLTDMKQVFTNAEYVNCRSGHVLSCNV 1430
Db 1388 SEVPDHOIKTFMDLAKIKSLNMGAYQLNEBLLSDIQLVFRNCGLYVNEGNEYIDAOC 1447
QY 1431 KTEQCLV 1437
Db 1448 QLRFVI 1454

RESULT 12

Q9NSL9 PRELIMINARY; PRT; 1427 AA.
AC Q9NSL9;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-WAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Hypothetical protein H20J04.2.
GN Name=H20J04.2; ORFNames=H20J04.2;
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
[1]
RC SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=9069613; PubMed=9851916;
RG WormBase Consortium;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
[2]
RC SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Dauphin S., Antoniou B., Gibson A.;
RT "The sequence of C. elegans fosmid H20J04.";
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
[3]
RC SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Waterston R.H.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
[4]
RC SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Waterston R.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
[5]
RC SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Waterston R.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
[6]
RC SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Waterston R.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
[7]
RC SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Waterston R.;
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
[8]
RC SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Waterston R.;
RL Submitted (OCT-2002) to the EMBL/GenBank/DBJ databases.
[9]
RC SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;

RA Waterston R.;
RN Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
[10]
RC SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RG WormBase Consortium;
RL Submitted (SEP-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC006663; AAF39888.2; -.
DR HSSP; Q03330; 1E6I.
DR WormBase; WBGene00019217; H20J04.2.
DR WormPep; H20J04.2; CE27187.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR000637; A+T hook.
DR InterPro; IPR001487; Bromodomain.
DR InterPro; IPR004022; DDT.
DR InterPro; IPR001965; Znf_PHD.
DR Pfam; PF02178; AT_hook; 1.
DR Pfam; PF00439; Bromodomain; 1.
DR Pfam; PF0628; PHD; 1.
DR PRINTS; PR00503; BROMODOMAIN.
DR SMART; SM00384; AT_hook; 1.
DR SMART; SM00297; BROMO; 1.
DR SMART; SM00571; DDT; 1.
DR SMART; SM00249; PHD; 1.
DR PROSITE; PS00633; BROMODOMAIN_1; 1.
DR PROSITE; PS00114; BROMODOMAIN_2; 1.
DR PROSITE; PS50827; DDT; 1.
DR PROSITE; PS01359; ZF_PHD_1; UNKNOWN_1.
DR PROSITE; PS50016; ZF_PHD_2; 1.
KW Hypothetical protein.
SQ SEQUENCE 1427 AA; 165227 MW; 401039F88C2BA4D3 CRC64;

Query Match 8.3%; Score 657.5; DB 2; Length 1427;
Best Local Similarity 20.2%; Pred. No. 7.2e-20;
Matches 331; Conservative 252; Mismatches 579; Indels 477; Gaps 61;

QY 3 PLLGRKPF---PLVNPPLGEEPPFTIPHTQEAFTREEYEARELYSERIWTCKSTGSSQ 59
Db 17 PLLHKQPEPQSVPGVTRDTVFYCKATKEVFLTHEEYFNRMILLNSTAWSCLTRKSN 76
QY 60 LTHKEAWEEEQEVAELKEEPFAWYKLVLEMVHH--NTASLEKLVDTAWLEIMTKYAVG 117
Db 77 LTYFEAIAASERAEQELC-NFPTALEIPITLVHKYTRNGRFDLVNDIYVHLKDRFFNN 135
QY 118 EECDFEVGKEKMLVKYIKHLEKVDDEATEKSDGACDSDSSDKENSQTAQDHQKE 177
Db 136 EEVAYAEKSRKFL-AKIVESNQIEB-----ONPSF--EGGSEL-----XE 172
QY 178 TVVKEDEGRRESINDRARRSPRKLPTSLKKGKRWAPPKFLPHKYDVKLQNEDKITS--- 234
Db 173 PIL-----PP---PDTFRYTLQLDPAVSEEE 196
QY 235 ----NVPADSLIRTERPPNKEIVRYFIRHNAAGTGENAPWVVEDELVKYSLPSKFSQ 290
Db 197 MYREGIAFDRLFRSKNIGSRQKIRLFLKNCCQMPDSE-----RYTLKQFLE 244
QY 291 FLLDPYKMYTLNPTSKRNTGSPDRKPSKSKTDNSSLSSPLNPKLWCHVHLKKSLSGSP 350
Db 245 KIDNLY-----WTDVNSGAEPICPQ----- 264
QY 351 LKVKNSKNSKSPHEEHEMMKMMSPNKLTNPHIPKGGPPAKPGKHSDKPKAKGRSKG 410
Db 265 -----TPALQGRAPNTLKAGDHERKE-----KKPKKEEDP-AAPPRPG 304
QY 411 ILNGQKSTGNSKSPKKGLTKPKTKMQMTLLDMAKGTQMTAPR-----NSGGPRTS 464
Db 305 -----RPP-----KTPQKALDKKEKQMRKRAEAGREAPKDDFDLTSSGA-MTL 348
QY 465 SKPHKHLPPAALHLIAYYKENKDRDKRSALSCVLSKTARLISSE-----DRARLPE 516
Db 349 SIP---VSPA-----KKKIRSGSSMSG--KKTKILEQOQELCFFTFSEARRLGI 394

Db 212 -----VTGEGWTDGLVFHGGTTP---SFOOSKRLK-- 237
QY 432 KTRKQKQMLDMAGKTQKMTAPRNSGGTPTSSKPKHKLPPAALHLIAYYIKENKDREK 491
Db 238 -----QNAERGEAGGG----- 249
QY 492 RSALSCVISKARLLSSDRARLPBELRSLVQKRYELLEKKWASMBEQRKEYLKKR 551
Db 250 -----AAASTALEAKAKQEBEAAQLRLVE-----QEALAKR 283
QY 552 EELKKLKEKAKERREKEMLERLEKO-----KRY-----EDOELTKGN-LPAFLVDTPEG 601
Db 284 ABEKKAL-----LAQQVVLAKRYNAVLEDQELPDORPLPPVRPRPL-- 326
QY 602 LPNTLFGDVAMVVEFLSCYSGLLLPDAQYP--ITAVSLMEAL--SADKGGFLYLNRLVL 658
Db 327 IAAHFSSFFILEYLNFPADLLAIRSKPPGGLTIHLERALLIREVNG--PLSDIFQVL 384
QY 659 LQTLLODEIAEDYGE-----LGMKLSEIPLTL 685
Db 385 LSAIFTLQOLEEENESVRYERIEGLAQKQTVPEQVRARDAALWCEKHYSTKLSEMPIDS 444
QY 686 HVSSELVRL-CLRRSDVQSESGSDTDNDKDSAFEDNE-----VQD--EFLEKLETSEFF 738
Db 445 MTSVELLRHLFLASGALVEEARHRYQNRGGYGDDPGLRLVRDYPHILRALNWSYV 504
QY 739 ELTSEBKQLTLALCHRIILMTYSVDHMETROQMS-----AELWKERLAVLKEENDK 790
Db 505 QLPVGGDIQLCLLIHQILTYSGVRLVVEERKEKARTASTVSQANRWARLLTVKA----- 560
QY 791 KRAEKOKRMEAKNKGKNGKVENGLGKTRKRIKVPFQVDTAEADMI SAVKSRLLAI 850
Db 561 -----GSLKNA-ARDDMKRELAGFEGELAAREE-----YRKLTD 594
QY 851 Q-----AKKEREIOREMVKLERQAEERIRKHKAABEKAQEGTAKAKLVNRTPIGT 905
Db 595 RLDEQCARTEADAQ-RQLKV---LQAESERLK-----EDFFDYQIY-----LGT 634
QY 906 DRNHRYWLFSDVEPGLFTEKGVHWDSDIDYRFNHHCKDHT---VSGDEDDYCPRSKXANI- 961
Db 635 DCRYNYWLF-BSLPLGLFVEH-----DRTVAGRLCDRPTHIPGLAACAPDQRKKYIT 686
QY 962 -----GKNASMTQH-GTATEVAVETTPKQGNLWELCDSQ----- 997
Db 687 RAIMRCAGTDGVVALLDAXAIDANAVSSSTGPAPVEBELLMCTAQPDPCVHTQHPG 746
QY 998 -----KELDELLNCLHPQIGRESQIKERLEKRYQDIITHSIHLARKP--NLGLKSC 1045
Db 747 TVGWGYATADELDALIRSLNARGAREKQURETLECE-RDLIVT-HIAKCPDLKLSVGEA 804
QY 1046 DGNQELLNFLRSLDIEVATRLONG-----GLGVVEETSEFEARV--- 1084
Db 805 D-----RAGALADIASRNOKRYDAPNFSHPGTEPNEILEAVFLENLLEAKITVG 856
QY 1085 -ISLEKLDGFCVIALQASVIKKFLQGFMA-----PKQKRKLQSEDSAKTEEVD 1134
Db 857 YLGVMEVRDRDKWREALEA-----HGYDAQTDRLVWGPKEKLAPKAEEGNGREQQDG 909
QY 1135 EKKQWVE-----AK 1144
Db 910 EQQVQKQENGPNDRVDSEQDKPVVAKLPGGGDLDSERDSLRLMSEAVQSLGRALLQ 969
QY 1145 VASALE-----XWKTARAEAOQTFSRMHWLLGML 1172
Db 970 VAAQCIDPKFLRHPFGKDKHDRAVNVQQKHQGLKNLVRVEVSLMQSTCYAQLFLHYNVL 1029
QY 1173 DACIKWMSAENARCKVCPKGGEDDKLILCDECNKAFHFLCRLRPALYEPDGEWCPAQ 1232
Db 1030 YDAIHSRSAREICCMICRKGDPPGLTLCDENCRACHYCYLKPKEVPAGDWTCKRCR 1089
QY 1233 P-----ATAPRNSRGRNVT-PESASEDSEDDSEEEEEEEEDYEVAGRLRPR 1285

Db 1090 PENFKEKPSANRKKKKMFSGSEDGAEBENDDGSDADYRAKRSKLMVKVAKVA-KRSKPK 1148
QY 1286 -----KTIR-----GKHSVIP 1296
Db 1149 AKAKPTIVTSTCRTSRSTAAATNGAYLALERWRRSEAOQCQPAVRHQEAQGGQDVSP 1208
QY 1297 PAARSORRRGKPKHSTRSQKAP-----PV-----DDAEVDLVLQTKRSRRROS 1342
Db 1209 PASKRARRA-----TNGRAKVSQPGTIDSDEEPGLGGGTGDDPLPLNSVALYT----- 1255
QY 1343 LELOKCEELHLKIVKRYFSWPRREPVTREADEYDVITHPMDFTQVQNKSCSGSYRSVQ 1402
Db 1256 -----LIDDLKHPNSWPNRPVSAKEVPDYAVIKSPMDFARIKSLNMGDYKINE 1307
QY 1403 EFLTDMKQVFTNAEVYN 1419
Db 1308 QMLSDVQLVFRNCOLYN 1324
RESULT 14
Q6PID9 PRELIMINARY; PRT; 892 AA.
AC Q6PID9
DT 05-JUL-2004 (Tremblrel. 27, Created)
DT 05-JUL-2004 (Tremblrel. 27, Last sequence update)
DT 05-JUL-2004 (Tremblrel. 27, Last annotation update)
DE BC065123 protein (Fragment).
GN Name=BC065123;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Brain;
RC MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Haieh F.,
RA Diachenko L., Marusina K., Farmer A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loughlan N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Blakesley R.W., Touchman J.W., Green E.D., Bouffard G.G.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences."
Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Brain;
RA Strausberg R.;
RL Submitted (JAN-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC065123; AAH65123.1; -.
DR GO; GO:003677; rDNA binding; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR001487; Bromodomain.
DR InterPro; IPR011011; FYVE PHD_Znf.
DR InterPro; IPR001965; Znf PHD.
DR Pfam; PF00439; Bromodomain; 1.
DR Pfam; PF00628; PHD; 1.
DR PRINTS; PRO0503; BROMODOMAIN.
DR SMART; SM00297; BROMO; 1.
DR SMART; SM00249; PHD; 1.
DR PROSITE; PS00633; BROMODOMAIN_1; 1.

DR	PROSITE; PS50014; BROMODOMAIN_2; 1.
DR	PROSITE; PS01359; ZF_PHD_1; 1.
DR	PROSITE; PS50016; ZF_PHD_2; 1.
FT	NON_TER 1
SQ	SEQUENCE 892 AA; 101830 MW; 95A6C798F37AA93F CRC64;
Query Match	
Best Local Similarity 24.5%; Score 640.5; DB 2; Length 892;	
Matches 216; Conservative 137; Mismatches 264; Indels 263; Gaps 28;	
QY	773 SAEIWKERLAVLKEENDKKAEBQKQKWEAKNKE--NGKVENGLGKTDRKKRIVK---P 827
DB	4 AAKIRRKKEKLKQEQKMKKEKQEKELKEDQORNSAAVPEEEREDFTSTENKNIHQKDL 63
QY	828 EPOVDTEAEADMI SAVK-SRRELLAIQAKKEREIOEREMKVKLRQA---EEERIRKHAA 882
DB	64 DDPVVTI EDDPGSHKESRRGKVGQTAVKQCIKEENNYCIKQPLSADAEALRQEQQ 123
QY	883 AEKAFQEGIIAKAKLVWRRTPIGTDRNHNRFSDVEPGLFISK---GWVHDSI-----D 934
DB	124 KERELLDKIQSAIACTNIFPLGRDRLRYRWIF-PSIPGLFIBEDYSGLTEDMLPRPS 182
QY	935 YRFNHCKOHTVSGDEDCPRS KANGLKNASMNTOHGTA TEVAVETTPKQONLWFLC 994
DB	183 FHNAQPRDPQVS-----IKTEESFLSESTSLDQ--GPFDDSVLLPKPVHKPNRCFY 234
QY	995 DSQKELDELNCLHPQIGRESQLKERLEKRYQDI-----HSIHLARKPNLGLK--- 1043
DB	235 SSCAQLDLDALNSGRHRESALKETLQEKSRICAQLAHFSEKTFHSDKPOADSKPVS 294
QY	1044 -----SCDGNQ-----ELLNFLRSLDIEVATRLQKGGIGY----- 1073
DB	295 SGRSSGACDISQMSABRQLELRLDFLLDIEDRIYQGTGLGAIKVTDRQVMSALENGRY 354
QY	1074 -----VEETSEFPARVISLEKL---KDFECVIALQAS----- 1103
DB	355 ELLSEESKENGVIKTVNDEEEMEQARVIVRDLRGIKETPTSTISTASTPQSVSNV 414
QY	1104 -----VIKKPLQGFMAPQKRRKLQSEDSAKTEEVDEKKWVEAKVASAL 1149
DB	415 VHYLALALFOEGIERFLK---AP-----LDGNDSGRSYK-----TVL 451
QY	1150 EKWKTAIRAEQTFSRMHVLLGMLDACIKWMSAENARCKVCPKKGEDDKLILCDECNKAF 1209
DB	452 DRWRESLLSSASLSQVFLHLSTLDRSVMSKSIILNARCKICRKKGAENVMVLCGCDGRH 511
QY	1210 HLCFLRPALYEPDPGEWQCACQP--ATARNRGRNYTE--ESASEDSDDSDSEDEEE 1265
DB	512 HTYCVRPKLKAVPDGDFWFCPECPKQRSRLSSRQPSLESDEMEEGMEDDDDEVD 571
QY	1266 EEEEEEDYEVA-----GLRLRPKRTIRGHSVIPPAARSGR-----RP 1305
DB	572 EEGQSEEEYEVEODESDSDDEALSPK--RGRPVRLPIKTKGRPGSPSPRSQRQDP 629
QY	1306 GKXPHSTRRSQPK-----APPVD-----DAEYDELV 1331
DB	630 GRYPERSQOSTPKNTAKSANKLRKRTSAPTETRSURVGRSRTRHSPSALQDVFFVLLS 689
QY	1332 -----LQTKRSSR-----RQS 1342
DB	690 PHSKRRGRKGADHTPEHSPSTFWNTFRVSTSRSSRLIPLNTAESLSLQHSERKGRKRQS 749
QY	1343 L-----ELQCEILHKIKYRFESWFPREPVITDEAEDYDVITHP 1383
DB	750 TESSPVFLNRRSGRQGVHELSAFQOLVVELVRHDDSDWFFPLKVLKSVIPDYDIHKP 809
QY	1384 MDFOTQNKSCGSYRSVQBFLLTDMKQVFTNAEVIYNCRGS 1423
DB	810 IALNIIEKYNKCEYKLASFIDDIELMFNCFEYNPRNT 849
RESULT 15	
Q6Y194	

Db 850 AFSDCLTIVEFLHSGKVLGDPADKDVPSLGLVQLQGLLQCGSLGSEVQDILLVRLKAAHL 909
QY 665 DEIAEDYGE-----LGMKLSIEPLTHSVSELVRLCLRRSDVQBESEGSDDTDKNKSAPE 720
Db 910 DPGFSPYCSQSLKILGEKVEIPLTRDNVSEILRCFLMAYG----- 949
QY 721 DNEVQDEFLEKLETSFEFFELTSEKLIQILTALCHRI-----LMTYSVQDHMETRQQMSAEL 776
Db 950 -----VFPALCDRLURTOPFOAQPOQQAUAFLVHLEUNGTLTLINEIDKLTSMSSYRKX 1006
QY 777 W-----KERLAVLKEENDKRAEKOKREMEAKNKENGKVENGLGKTDRKRIIVFEPQVDT 834
Db 1007 WIVEGRLRLK-----TVLAKRTGRSEVENEGP-----EELGRRSSRIMEETSGWEEB 1057
QY 835 ADMISAVKSRLLAIQAKKEBIEQEREMKV-KLERQABEEBIRKHAEEAKAFQEGIAK 893
Db 1058 EESIAAAMPGR-----GRDGEVDATASSIPELERQIEKLSKR-----QLFFRKLHL 1106
QY 894 AKLVMRRTPIGTDRNHRVWLFSDVPGLPFIE----- 925
Db 1107 SSQMLRAVSLGQDRYRRRYVVL-PYLAGIFVETGTEGNLVPVEVVKETDSLKVAHASLN 1165
QY 926 ----- 925
Db 1166 PALFSKMLAGSNTTASSPARAGRPRKTKPGSMQPHLKSPVRGQDSQEQPAQLOPEA 1225
QY 926 -----KGVW-----HD-----SIDYRPNHCKDHT----- 945
Db 1226 QLHAPAQPOQLQLQSHKGFLEGBGSLGSLQSHDLQSQAFSLWNLQTSQSHSLSS 1285
QY 946 -----VSGD 949
Db 1286 SVLTPDSSPKGLDPAPSPPBPDEABSSPDQALWFNISAQMPNCAAPTTPPAVSD 1345
QY 950 EDYCPRSKXANLGKXASMN----- 968
Db 1346 Q-----PTPSPQLASSKPMNRPANPCSPVQSSPTPLAGLAKPRAGDPGMPQSPGTGL 1402
QY 969 -----TOHGTATEVAVETTTPOGQNLWFLCDQSKELDBLLNCLHPOGIR 1013
Db 1403 QPKRRGRPPSKFKQMEQRYLTQLTAQVPVPEMCGWMIWIRPEW-LDAMLKALHPRGIR 1461
QY 1014 ESQLEKRLKRYODIHSIHL-----ARKPNLGLKSCDGNQF-LNLFRLSLDIEVATRLQKG 1070
Db 1462 EXALHKLHLN-IRDFLEQVCLRPSADPIPEPQLPAFQEGISWSPKE-----KTYETD 1514
QY 1071 LGVVESTSEFARVIS-----LEKLKDFGECV----- 1097
Db 1515 LAVLQWVELEQVIMSDIQRGWTCPSPDSTREDLAYCEHLSDSQEDITWGRGREGIA 1574
QY 1098 -----IALQASVKKFLQGFQWAPKQKRLKLOSDSQAkteevDEEKWVE 1141
Db 1575 POKTTNPLDLAVMLAALAEQNVRRYRLREPLW-----THEVVLEKALLS 1620
QY 1142 -----EAKVASALEKWKTAIREAQTFSRMVLGLMDACIKWDSAEANARCK 1188
Db 1621 TNGAPEGTTTSEIYTPRIRVWOTLERCSRSAQVCLGLQLESIWKSXVNVKTL 1680
QY 1189 VCPKKGEDKILDCBNKAFHLCLRLPALYEVDPGEWQCPACQAPATARRNSGRNRYTEE 1248
Db 1681 VCRKGNDFLLJCDGDRGCHYCHRPKWEAVPEGDWFTVC----- 1723
QY 1249 SASDESDDESDEEBEEDYEV-----AGLRLRPKRTIRGKHSVIPPAARSRRP 1305
Db 1724 -LAQQVEGFTQKPPKRGKRGKSYLNFSEGDGRRRRVLLRGRES-----PAAGPRYSEE 1780
QY 1306 GKPKHSTRSQKAPVDDAEDVDELVTQKRSRRQSLSQKEETLHKVYKRSWPPR 1365
Db 1781 GLSPSKRRR-----LSMRNHSDLTFCEIILMEWESHDAWPL 1819
QY 1366 EPVTRDEADYDVITHPMDFTQVQNKCSGYSRVSQEFITDMQVFTNAEYVNCRGSHV 1425
Db 1820 EPVNPRLVSGYRRIKQNPMDFTMRERLLRGGYTSSEEPADALLVFDNCQTFNEDDSEV 1879

RESULT 17
BA2B HUMAN
ID BA2B HUMAN STANDARD; PRT; 1972 AA.
AC Q9UIF8; Q96EA1; Q96SQ8; Q9P252; Q9Y4N8;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Bromodomain adjacent to zinc finger domain 2B (hwalp4).
GN Name=BAZ2B; Synonyms=KIAA1476;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Testis;
RX MEDLINE=20130112; PubMed=10662543; DOI=10.1006/geno.1999.6071;
RA Jones M.H., Hamana N., Nezu J., Shimane M.;
RT "A novel family of bromodomain genes.";
RL Genomics 63:40-45(2000).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 3).
RA Ohara O., Nagase T., Kikuno R.;
RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 753-1972 FROM N.A.
RX MEDLINE=20277482; PubMed=10819331;
RA Nagase T., Kikuno R., Ishikawa K.-I., Hirosewa M., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XVII.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 7:143-150(2000).
RN [4]
RP SEQUENCE OF 1-776 FROM N.A. (ISOFORM 1).
RC TISSUE=Melanoma;
RA Ansong W., Winkler U., Mewes H.-W., Weil B., Wiemann S.;
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE OF 1-799 FROM N.A. (ISOFORM 2).
RC TISSUE=Skeletal muscle;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.P., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield J.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,
RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [6]
RP SEQUENCE OF 427-1215 FROM N.A. (ISOFORM 3).
RX PubMed=14702039; DOI=10.1038/ng1285;
RA Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,
RA Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,
RA Sekine M., Ohyashiki M., Nishi T., Shibahara T., Tanaka T., Ishii S.,
RA Yamamoto J.-I., Saito K., Kawai Y., Isono Y., Nakamura Y.,
RA Nagahara K., Murakami K., Yasuda T., Watanabe T., Wagatsuma M.,
RA Shiratori A., Sudo H., Hosoiri T., Kaku Y., Kodaira H., Kondo H.,
RA Sugawara M., Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E.,

RA Omura Y., Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa M.,
RA Yamazaki M., Ninomiya K., Ishibashi T., Yamashita H., Murakawa K.,
RA Fujimori K., Tanai H., Kimata M., Watanabe M., Hiraoka S., Chiba Y.,
RA Ishida S., Ono Y., Takiguchi S., Watanabe S., Yosida M., Hotuta T.,
RA Kusano J., Kanehori K., Takahashi-Fujii A., Hara H., Tanase T.-O.,
RA Nomura Y., Togiya S., Konai F., Hara R., Takeuchi K., Arita M.,
RA Inose N., Musashino K., Yuuki H., Ohnima A., Sasaki N., Aotsuka S.,
RA Yoshikawa Y., Matsunawa H., Ichihara T., Shiohata N., Sano S.,
RA Moriya S., Momiya H., Sato H., Takami S., Terashima Y., Suzuki O.,
RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,
RA Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,
RA Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,
RA Fujimori Y., Koniyama M., Tashiro H., Tanigami A., Fujiwara T.,
RA Ono T., Yamada K., Fujii Y., Ozaki K., Hirao M., Ohmori Y.,
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Okitani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,
RA Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,
RA Togaashi T., Oyama M., Hata H., Watanabe M., Komatsu T.,
RA Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,
RA Okumura K., Nagase T., Nomura N., Kikuchi H., Masuho Y., Yamashita R.,
RA Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.,
RT "Complete sequencing and characterization of 21,243 full-length human
RT cDNAs.";
RL Nat. Genet. 36:40-45(2004).
RN [7].
RP SEQUENCE OF 1524-1972 FROM N.A.
RC TISSUE=Testis;
RL Poustka A., Klein M., Mewes H.-W., Gassenhuber J., Wiemann S.;
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: May play a role in transcriptional regulation
CC -!- interacting with ISWI.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=3;
CC Comment=Experimental confirmation may be lacking for some
CC isoforms;
CC Name=1;
CC IsoId=Q9UIF8-1; Sequence=Displayed;
CC Name=2;
CC IsoId=Q9UIF8-2; Sequence=VSP_000553;
CC Note=Inferred from Ref.4;
CC Name=3;
CC IsoId=Q9UIF8-3; Sequence=VSP_000554;
CC Note=Inferred from Ref.5;
CC -!- TISSUE SPECIFICITY: Expressed at varying levels in several
CC tissues, whereas a smaller transcript was expressed specifically
CC in testis.
CC -!- SIMILARITY: Belongs to the WAL family.
CC -!- SIMILARITY: Contains 1 bromodomain.
CC -!- SIMILARITY: Contains 1 DDT domain.
CC -!- SIMILARITY: Contains 1 methyl-CpG-binding (MBD) domain.
CC -!- SIMILARITY: Contains 1 PHD-type zinc finger.
CC -!- CAUTION: Ref.5 sequence differs from that shown due to a
CC frameshift in position 731.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB032255; BAA89212.1; -;
DR EMBL; AB040909; BAA96000.2; ALT_INIT.
DR EMBL; AL834381; CAD30044.1; ALT_INIT.
DR EMBL; BC012576; AAH12576.1; ALT_FRAME.
DR EMBL; AK027612; BAB55231.1; ALT_INIT.
DR EMBL; AL080173; CAB45759.1; -;
DR FIC; T12495; T12495.
DR HSSP; OSUIG0; 1F62.
DR GenSeq; HGNC:963; BAZ2B.
DR MIM; 605683; -;
DR InterPro; IPR001487; Bromodomain.

DR InterPro; IPR004022; DDT.
DR InterPro; IPR001739; Methyl-CpG_bind.
DR InterPro; IPR001965; Znf_PHD.
DR Pfam; PF00439; Bromodomain_1.
DR Pfam; PF02791; DDT; 1.
DR Pfam; PF01429; MBD; 1.
DR Pfam; PF00628; PHD; 1.
DR PRINTS; PR00503; BROMODOMAIN.
DR PROSITE; PS00633; BROMODOMAIN_1; FALSE_NEG.
DR PROSITE; PS00014; BROMODOMAIN_2; 1.
DR PROSITE; PS00827; DDT; 1.
DR PROSITE; PS00827; MBD; 1.
DR PROSITE; PS01359; ZF_PHD_1; FALSE_NEG.
DR PROSITE; PS00016; ZF_PHD_2; 1.
KW Alternative splicing; Bromodomain; Coiled coil; DNA-binding;
KW Nuclear protein; Transcription regulation; Zinc-finger.
FT DOMAIN 543 614 MBD.
FT DOMAIN 891 956 DDT.
FT ZN_FING 1735 1785 PHD-type.
FT DOMAIN 1881 1951 Bromodomain.
FT DOMAIN 45 69 Ser-rich.
FT DOMAIN 73 79 Poly-Glu.
FT DOMAIN 399 470 Asp/Glu-rich (acidic).
FT DOMAIN 641 672 Arg-rich.
FT DOMAIN 687 865 Coiled coil (potential).
FT DOMAIN 706 875 Lys-rich.
FT DOMAIN 1100 1143 Asp-rich.
FT DOMAIN 1138 1179 Coiled coil (potential).
FT VARSPIC 437 534 Missing (in isoform 2).
FT VARSPIC 593 626 Missing (in isoform 3).
FT CONFLICT 137 137 S -> F (in Ref. 1).
FT CONFLICT 226 226 S -> L (in Ref. 1).
FT CONFLICT 627 627 G -> E (in Ref. 1).
FT CONFLICT 722 722 E -> K (in Ref. 5).
FT CONFLICT 790 799 RIKKKKKKKK -> QKKKKKKKKK (in Ref. 5).
FT CONFLICT 1128 1128 K -> Q (in Ref. 1).
FT CONFLICT 1183 1183 Q -> P (in Ref. 1).
FT CONFLICT 1195 1195 Q -> R (in Ref. 1).
FT CONFLICT 1453 1453 L -> S (in Ref. 3).
FT CONFLICT 1838 1838 K -> Q (in Ref. 1).
SQ SEQUENCE 1972 AA; 220709 MW; D4F105C927B91569 CRC64;
Query Match 6.7%; Score 536; DB 1; Length 1972;
Best Local Similarity 18.1%; Pred. No. 1.5e-14;
Matches 355; Conservative 258; Mismatches 596; Indels 750; Gaps 68;
QY 8 KPFPVLNPLPGEPP-FTIPH-----TQEAFTREEYEARELYSERIWTCKST 55
DB 189 KFLSLVNQAKKTYMKLVIPSPDLKAGNKNTSESSSLTSELSKREQY-----KQA 241
QY 56 GGSQLTKEAWBEEQVABLLK-----EEFPWYKLVLEMVHNN-----TA 97
DB 242 FPSQLKQBSKSLKVKVIAALSNNPRATSSSPA-HPKQTLNHNPNPFLNALLGNHPNG 300
QY 98 SLEKLVDTAWLIMTKYAVGECDEEV-----124
DB 301 VIQSVIQEAPLALTTKTKMQSKINENIAAASSTPFPSPVNLSTSGRRTPGQNPVMPAS 360
QY 125 -----GKEML--KVKIVKI-----HPL-----EKVDEE 146
DB 361 PILHSGKEKAVNNVNPVKTOHSHPAKSLVQFQGTSDIPSSKSDSEDEEEDDE 420
QY 147 ATEKXSDGACDPSKDKENSQIAQHQKTVVKEDEGRRSINDRARRSRKPLTSLK 206
DB 421 EEDEDEDEDDSDSQESDSNSSESTGESEEDDDKQDESDS-----T 467
QY 207 KGERKWAPEKFLPHKYDVVKLQNEKLIENV--PADSLITERPPNKEIVYRIFRHNALRA 264
DB 468 EGE-----KTMKL-----NKTTSVKPSMSLTGHSTPRNLIAK----- 503
QY 265 GTGENAPWVEDELVKVYSLPSKFSDFLLDPKYKMTNPSTKRKNTGSPDRKPSKSK-T 323

SEQUENCE OF 1038-1878 FROM N.A.
 RP TISSUE=Lymph;
 RC MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
 RA Klausner R.D., Collins F.S., Wagner L., Shemmen C.M., Schuler G.D.,
 RA Atschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan K., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Donald M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters K.J., Abramson R.D., Mullaly S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 CC -!- FUNCTION: May play a role in transcriptional regulation
 CC interacting with ISWI. May serve a specific role in maintaining or
 CC altering the chromatin structure of the rDNA locus (By
 CC similarity).
 CC -!- SUBUNIT: Together with ISWI/SNP2h, it forms a complex termed NORC
 CC (nucleolar remodeling complex).
 CC -!- SUBCELLULAR LOCATION: Nuclear. Colocalizes with the basal RNA
 CC polymerase I transcription factor UBF in the nucleolus.
 CC -!- TISSUE SPECIFICITY: Expressed at moderate levels in most tissues
 CC analyzed, including heart, brain, placenta, lung, skeletal muscle,
 CC kidney and pancreas.
 CC -!- SIMILARITY: Belongs to the WAL family.
 CC -!- SIMILARITY: Contains 4 A.T hook DNA-binding repeats.
 CC -!- SIMILARITY: Contains 1 bromodomain.
 CC -!- SIMILARITY: Contains 1 DDT domain.
 CC -!- SIMILARITY: Contains 1 methyl-CpG-binding (MBD) domain.
 CC -!- SIMILARITY: Contains 1 PHD-type zinc finger.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC -----
 CC EMBL; AB032254; BAA89211.1; -;
 CC EMBL; AF000422; AAB60864.1; -;
 CC EMBL; AB002312; BAA20773.1; -;
 CC EMBL; BC008965; BAA08965.2; -;
 CC HSSP; Q9UIG0; 1F62.
 CC GenSeq; HGNC:962; BAZ2A.
 CC MIM; 605682; -;
 CC GO; GO:0005731; C:nucleolus organizer complex; NAS.
 CC GO; GO:0003677; P:DNA binding; NAS.
 CC GO; GO:0030528; P:transcription regulator activity; NAS.
 CC GO; GO:0006338; P:chromatin remodeling; NAS.
 CC GO; GO:0006355; P:regulation of transcription, DNA-dependent; NAS.
 CC InterPro; IPR000637; A.T hook.
 CC InterPro; IPR001487; Bromodomain.
 CC InterPro; IPR004022; DDT.
 CC InterPro; IPR001739; Methyl-CpG_bind.
 CC InterPro; IPR001965; Znf_PHD.
 CC Pfam; PF02178; AT_hook; 4.
 CC Pfam; PF00439; Bromodomain; 1.
 CC Pfam; PF02791; DDT; 1.
 CC Pfam; PF01429; MBD; 1.
 CC Pfam; PF00628; PHD; 1.
 CC PRINTS; PR00503; BROMODOMAIN.
 CC SMART; SM00384; AT_hook; 4.

DR SMART; SM00297; BROMO; 1.
 DR SMART; SM00571; DDT; 1.
 DR SMART; SM00391; MBD; 1.
 DR SMART; SM00249; PHD; 1.
 DR PROSITE; PS00633; BROMODOMAIN_1; FALSE_NEG.
 DR PROSITE; PS50014; BROMODOMAIN_2; 1.
 DR PROSITE; PS50827; DDT; 1.
 DR PROSITE; PS50982; MBD; 1.
 DR PROSITE; PS01359; ZF_PHD_1; FALSE_NEG.
 DR PROSITE; PS0016; ZF_PHD_2; 1.
 KW Bromodomain; Coiled coil; DNA-binding; Nuclear protein; Repeat;
 KW Transcription regulation; Zinc-finger.
 KW DOMAIN 519 590 MBD.
 FT DNA_BIND 622 634 A.T hook 1.
 FT DOMAIN 821 886 DDT.
 FT DNA_BIND 1159 1171 A.T hook 3.
 FT DNA_BIND 1377 1389 A.T hook 4.
 FT ZN_FING 1649 1699 PHD-type.
 FT DOMAIN 1783 1853 Bromodomain.
 FT DOMAIN 633 772 Lys-rich.
 FT DOMAIN 666 765 Coiled coil (Potential).
 FT DOMAIN 1185 1250 Glu-rich.
 FT DOMAIN 1263 1384 Pro-rich.
 FT DOMAIN 1732 1735 Poly-Arg.
 FT CONFLICT 574 574 L -> V (in Ref. 2).
 FT CONFLICT 700 700 L -> Q (in Ref. 2 and 3).
 FT CONFLICT 720 720 H -> Q (in Ref. 2 and 3).
 FT CONFLICT 727 738 SNAKEKGTQ -> KIKKKKKKKK (in Ref. 2).
 FT CONFLICT 785 785 K -> R (in Ref. 3).
 FT CONFLICT 951 951 P -> L (in Ref. 3).
 FT CONFLICT 1005 1006 GR -> EG (in Ref. 3).
 FT CONFLICT 1163 1163 G -> S (in Ref. 1).
 FT CONFLICT 1166 1166 R -> L (in Ref. 1).
 FT CONFLICT 1172 1172 S -> F (in Ref. 1).
 FT CONFLICT 1178 1178 L -> F (in Ref. 1).
 FT CONFLICT 1202 1202 A -> V (in Ref. 1).
 FT CONFLICT 1292 1292 P -> L (in Ref. 1).
 FT CONFLICT 1295 1295 L -> F (in Ref. 1).
 FT CONFLICT 1313 1313 P -> L (in Ref. 1).
 FT CONFLICT 1407 1410 Missing (in Ref. 4).
 FT CONFLICT 1416 1416 R -> P (in Ref. 1).
 FT CONFLICT 1541 1541 R -> P (in Ref. 1).
 FT CONFLICT 1571 1571 E -> K (in Ref. 1).
 FT CONFLICT 1616 1616 V -> I (in Ref. 1).
 FT CONFLICT 1622 1622 E -> Q (in Ref. 1).
 FT CONFLICT 1629 1629 Q -> H (in Ref. 1).
 FT CONFLICT 1636 1636 Q -> H (in Ref. 1).
 FT CONFLICT 1739 1739 R -> K (in Ref. 1).
 FT CONFLICT 1754 1754 G -> R (in Ref. 1).
 SQ SEQUENCE 1878 AA; 208480 MW; 400970CA68234317 CRC64;
 Query Match 6.7%; Score 534.5; DB 1; Length 1878;
 Best Local Similarity 19.4%; Pred. No. 1.7e-14;
 Matches 265; Conservative 176; Mismatches 440; Indels 483; Gaps 41;
 QY 415 OKSTGNSKPKGLTKPTKTKMTLDWAKGTQ---KMTAPRNSGGPTTSKPKHKL 471
 Db 619 QAITGRGRPR---NTEKATREVKVKGRRPKVKITEL-----LNKTDRPLKKL 669
 QY 472 PPAALHLIAYYKENKDRKRSALSCVISKTARLLSSEDRARLPEELSLVQKRYELLEH 531
 Db 670 EA-----QETLNEEDKAKIAKSKKQKQVQGECLTTIQGQARNKQKETSLSKH 720
 QY 532 KURWASMSSEQRKYLKKRBEELKKLKEKAKER---REKE----- 569
 Db 721 KE--AKKSKAEKKGKTKQEKLEKVKREKKEKVKKEEVTAKAKACKADKTLATOR 778
 QY 570 MLERLEKQ-----KRYEDQELTG-KNLPAFLVDTPEG--LPNTLFGDVAMVVEFLS 618
 Db 779 RLEERKQKQMILEEMKKTPTEDMCLTDHQPPLPFSRV---PGLTLPSGAFSDCLTIVEFLH 835
 QY 619 CYSGLLLPDAQYPITAVS-LMEALSADKGGFLYLNRLVILLQTLQDIAEDYGE---- 673

Db	836	SFGKVLGFDPAKVPVSLVQLQEGLLCQDSLGELVQLVRLKKAALHDPGFPFSYCSLSKI	895
Qy	674	LGMLSEIPLTLHVSSEVLRLCLRRSDVOESEGSDTDNDKSDSAFEDNEVQDEPLEKLE	733
Db	896	LGEKVSIEPLTRDNDVSEILRCFLMAYG-----VEPALCDRLR	932
Qy	734	TSEFPELTSEKLIQLITLALCHRI-----LMTYSVQDHMETROQMSAELW--KERLAVLKEE	787
Db	933	TQPPQAQPPQQAVALAPFVHLNGSTLIINEIDKTLESMSYRKNKWIIVGRLRLK--	990
Qy	788	NDKRAEKOKKEMEKAKNGKGVNGLGKTDKRRKRIVKFEPQVDTEAEDMISAVKSRL	847
Db	991	--TVLAKTGRSEVM-----GRPECLGRRRSSRIMEETSGMBEEEEESIAAVPGR--	1042
Qy	848	LAIQAKKEREIOEREMKV-KLERQAEERIRKHAABKAFQEGIAKALVMRRTPIGTD	906
Db	1043	---GRRDGEVDATASSIPELERQIEKLSK-----QLFFRKLLHSSQMLRAVSLQD	1092
Qy	907	RHNRYWLFSDVPGLFTE-----KGWVHDSIDYRF-----	937
Db	1093	RYRRRYWL-PYLAFIFVEGTGNLVPPEVIKKTDSLKVAHAASINPALFSMKMELAGS	1151
Qy	938	-----NHHCKDHTVSGDEDYCPRSK-----	957
Db	1152	NTTASSPARAGRPRKTPGSGMQPRHLKS-PVRGQDSQOPQALQPEALQALHAPQPPQL	1210
Qy	958	-----KANLGN-----ASNTHGTGATEVAVETTPKQG--	987
Db	1211	QLQLQSHKGFLQEGSPLSGSQHSDLSQSAFLWSLQTSQSHSLSSSVLTPDSSPGKL	1270
Qy	988	-----QNWLFLCDSQ-----	997
Db	1271	DPAPSQPPEPEDEAESSPPQALWFNISQAMPNCAAPTTPPAVSEDOPTSPQOLASS	1330
Qy	998	-----	997
Db	1331	KPMNRPSAANPCSPVQFSSTPLAGLAKRRAGDQGMPCQSTGLQOPKRRGRPPSPKFKQ	1390
Qy	998	-----KELDELNCLHLPQGIRESQKLERKRYQDI1	1029
Db	1391	MEQRYLTQLTAQVPVPENCSGWWIRDPEDMLDAMLKALHPRGIREKALHKLNK-HRDFL	1449
Qy	1030	HSIHL--ARKNLGLKSCDGNQE-LLNFRSLDLIEVATRLQKGLGVVEETSEFFARVIS	1086
Db	1450	QEVCLRPSADPIFEPLQPLAFQEGIMSWSPKE-----KTYETDLAVLQWVEELQORVIM	1503
Qy	1087	-----LEKLDQFGECV-----	1097
Db	1504	SDLQIRGWTCSPDSTREDLAYCEHLSQSDITWGRGREGCLAPQRKTTNPLDLAVMRL	1563
Qy	1098	IALQASVIKKFLQGFMAPKQKRKLQSDSAKTEEVDEEKKMVE-----BAK	1144
Db	1564	AALEQNVERYLREPLW-----THEVLEKALLSTPNGAPEGTTEIWEIS	1609
Qy	1145	VASALEKWKTAIREAQTSRMHVLGLMDACIKWDMSAENACKVCPKKGEDDKLILCDE	1204
Db	1610	ITPRIRVWRQTLERCSAAQVCLCGQLERSIAWEKSVNKVTCVLCVRKGDNDFFLLCDG	1669
Qy	1205	CNKAHFLCLRALYVDPGEWQCPACQAPATARRNSGRNRYTEESASEDSDESDDEEE	1264
Db	1670	CDRGCHYIYCHRPMEAVPEGMFCTVC-----LAQQVEGEFTQKPGF	1711
Qy	1265	EEEEEEEDYEV---AGLRLPRKTIIRTKGHKSVIPPAARSGRRPKGKHSTRRSPKAPP	1321
Db	1712	PKRGQKXSGYSLNSEGDRRRVLLRGRES--PAAQPRYSEEGLSFSKRR-----	1762
Qy	1322	VDDAEVDELVLQTKRSRRQSLELOKCEILHKIVKRYFSWPFREPVTRDEADYDVIT	1381
Db	1763	-----LSMNHHSDLATFCEIILMEMESHDAAMPFLFEPVNPRLVSGYRRIK	1808
Qy	1382	HPMDFQTVQNKSCSGSYSVQSEFLDMKOVFTNAEVYTCRGS HV	1425

1809 NPMDFSTMRLLRGGVTSSEFAADALLVFDNCQTFNDDSEV 1852

RESULT 19

BAZA MOUSE

ID BAZA MOUSE STANDARD; PRT; 1850 AA.

AC Q91VE5;

DT 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 05-JUL-2004 (Rel. 44, Last annotation update)

DE Bromodomain adjacent to zinc finger domain 2A (Transcription

DE termination factor-1 interacting protein 5) (TTF-I interacting protein

DE 5) (Tip5).

GN Name=Baz2a; Synonyms=Tip5;

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=21423468; PubMed=11532953; DOI=10.1093/emboj/20.17.4892;

RA Strohnner R., Nemeth A., Jansa P., Hoffmann-Rohrer U., Santoro R.,

RA Laengst G., Grummt I.;

RT "NORC-a novel member of mammalian ISWI-containing chromatin remodeling

RT machines";

RL EMBO J. 20:4892-4900(2001).

CC -!- FUNCTION: May play a role in transcriptional regulation

CC interacting with ISWI. May serve a specific role in maintaining or

CC altering the chromatin structure of the rDNA locus.

CC -!- SUBUNIT: Together with ISWI/SNF2h, it forms a complex termed NORC

CC (nucleolar remodeling complex).

CC -!- SUBCELLULAR LOCATION: Nuclear. Colocalizes with the basal RNA

CC polymerase I transcription factor USF in the nucleolus.

CC -!- SIMILARITY: Belongs to the WAL family.

CC -!- SIMILARITY: Contains 4 A.T hook DNA-binding repeats.

CC -!- SIMILARITY: Contains 1 bromodomain.

CC -!- SIMILARITY: Contains 1 DDT domain.

CC -!- SIMILARITY: Contains 1 methyl-CpG-binding (MBD) domain.

CC -!- SIMILARITY: Contains 1 PHD-type zinc finger.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

CC EMBL; AJ309544; CAC69992.1; --.

DR HSPF; Q9UIG0; 1F62.

DR MGD; MGI:2151152; Baz2a.

DR GO; GO:0005731; C:nucleolus organizer complex; ISS.

DR GO; GO:0003677; F:DNA binding; ISS.

DR GO; GO:0030528; F:transcription regulator activity; ISS.

DR GO; GO:0006338; P:chromatin remodeling; ISS.

DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; ISS.

DR InterPro; IPR000637; A+T hook.

DR InterPro; IPR001487; Bromodomain.

DR InterPro; IPR004022; DDT.

DR InterPro; IPR001739; Methyl-CpG_bind.

DR InterPro; IPR001965; Znf.PHD.

DR Pfam; PF02178; AT_hook; 4.

DR Pfam; PF00439; Bromodomain; 1.

DR Pfam; PF02791; DDT; 1.

DR Pfam; PF01429; MBD; 1.

DR Pfam; PF00628; PHD; 1.

DR PRINTS; PR00503; BROMODOMAIN.

DR SMART; SM00384; AT_hook; 4.

DR SMART; SM00297; BROMO; 1.

DR SMART; SM00571; DDT; 1.

DR SMART; SM00391; MBD; 1.

DR SMART; SM00249; PHD; 1.

DR PROSITE; PS00633; BROMODOMAIN_1; FALSE_NEG.


```
QY 979 VETTTTQKQNLWFLCDSQKELDELLNCLHPOGIIRESQKLEKRYQYDIHSIHL----- 1034
Db 1400 EPQVVPREFRYGKWKINDIEELNELIKALNPGVRERLURQSL-----LSLSAESVNLTPH 1456
QY 1035 -----ARKPNLGLKSCDGNQELLNFLRS-----DLIE-----VATRLQKGGGLGYVE 1075
Db 1457 HVSHPRAPPPNGYIEPEAWNAWNSIARRVEVALLDQIEAMEDKVASASMQVKQWMPQ 1516
QY 1076 ETSEFEARVISLEKUKDFGECVIALQASVIKFLQGFMAPKQKRKL-----QSEDSAKTE 1131
Db 1517 REGDSENGVVEDVITTEMLRERILGLEAATERLYLPPGLGINTTEAQMAMVIAQOESHQN 1576
QY 1132 EYDE-----EKKVBEAKVASALEKNTAIREAQTFSRMHVLGMLDCACTKWDMSAENARC 1187
Db 1577 NVSNLNSGNSAEDENLPKGLLWRDAVERSTTAQLSMALYLVESCVAMDKSIWKANC 1636
QY 1188 KVCPPKGGEDKILDCENKAFHLCFLRPALYVEPDGEWQCPACQ-PATARN----- 1239
Db 1637 QFCQSGESEDKLLDCGDRGYHTYCFKPRMDKIPDGDWYCFECKNKATGDRKCIVCGL 1696
QY 1240 -----SGRNYTEE----- 1248
Db 1697 RPPLGKMWYCLCPRAYHQDCYIIPMLKYPRGWYQNCVAKAPPKKPKQPKERTTN 1756
QY 1249 -----SASEDESD----- 1260
Db 1757 NSSQLLNSLNSNQNSLNSHEDIATPLSATSAATVEQAQSDYVASGEAAQYSASE 1816
QY 1261 -----EEEEEEEEEDYEVAGLRPRKTIRGKHSVIP----- 1296
Db 1817 GYTHAQSTSECPDQESPAEVDQIEASSESPVPTAGGEGFYSPSSNMNNAAGSSTVL 1876
QY 1297 -----PAASGRPPGKKHSTRSQPKAPPVDDA----- 1325
Db 1877 DSDRTFGGDGEGSGGDSSEYOQRPSPVPKDSLCLAGSSSSSSSSISDHQKERSKER 1936
QY 1326 -EVDLVLQTKSSRSQSLQKCEILHKIVKYSMPFREPVTREDAEDYDVITHPM 1384
Db 1937 DKAERAKQEKATKRLKELAVCKTILEEMELHEDSWPFLPVNTKOPFTYRKRVIKSPM 1996
QY 1385 DQTVQNKSCGSYSRVSQEFLLDMKQVFTNARVYN 1419
Db 1997 DLSTIKRLQDLVYSREDFIADVRQIFDNCVEFN 2031
ID O45075 PRELIMINARY; PRT; 1370 AA.
AC O45075;
DT 01-JUN-1998 (TReMBLrel. 06, Created)
DT 01-OCT-2002 (TReMBLrel. 22, Last sequence update)
DT 01-MAR-2004 (TReMBLrel. 26, Last annotation update)
DE Hypothetical protein F53H1.4.
GN Name=F53H1.4; ORFNames=F53H1.4;
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Feloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=99069613; PubMed=9851916;
RG WormBase Consortium;
RT "Genome sequence of the nematode C. elegans: a platform for
RL investigating biology. The C. elegans Sequencing Consortium.";
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Beck C., Kramer J., Keppler D.;
RT "The sequence of C. elegans cosmid F53H1.4";
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
```

```
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Waterston R.;
RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RG WormBase Consortium;
RL Submitted (SEP-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF045641; AAC02578.2; -.
DR PIR; G88637; G88637.
DR HSSP; Q13263; 1FP0.
DR WormBase; WBGene00018778; F53H1.4.
DR WormPep; F53H1.4a; CE31539.
DR GO; GO:0003677; P:DNA binding; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR001965; Znf_PHD.
DR Pfam; PF00628; PHD; 1.
DR SMART; SM00249; PHD; 1.
DR PROSITE; PS01359; ZF_PHD_1; UNKNOWN_1.
DR PROSITE; PS00016; ZF_PHD_2; 1.
KW Hypothetical protein.
SQ SEQUENCE 1370 AA; 158292 MW; 1E508414B13CC534 CRC64;

Query Match 5.7%; Score 455.5; DB 2; Length 1370;
Best Local Similarity 19.9%; Pred. No. 2.7e-11;
Matches 317; Conservative 195; Mismatches 531; Indels 547; Gaps 59;

QY 71 EVAELLKEFPAPWYEKLVLEMHVHNTASLEKLVDTAMLEIMTKYAVGECDFEVGKEXML 130
Db 74 EIEDLIDEKFPAG-EKLMSSGKEYTVVSSEKRGGLTYTMEDGTGKIGHR---DLRRKKGL 129
QY 131 KVKIVK---IHPLEKVDSEATEKKSDGACDSP-----SSDKENSQIAQ----- 171
Db 130 SVEEIKTIAIEDAEFVDEKQWKVREELAEPIREIKKYAPIFSANRKSPTKTAQLSVAA 189
QY 172 --DHQKQKTVVKEDEGRRESINDARRSPKLPSTLKKGERKQWAPPKFLPHKYDYKLQNE 229
Db 190 EADSDVQEVSMDSAEAGAAANKSAMKTPRGAP-----RASGGFVLSRR---RLQEK 239
QY 230 DKTIISNVPADSLIRTPPNKEIVRYFIRHNALRAGTGENAPWVVEDELVKYSLSPKFS 289
Db 240 QKEDKEKEKLEKKQKEQ-----EKKQKKEEKAKK--LKEK-E 277
QY 290 DFLDLPYKYMTLNPSTKTKNTGSPDRKPKSKSTDNSSLSPLNPKLWCHVHLKSLSGS 349
Db 278 EKLKEKEKAARAKAEKKEKNGTMDK-----FLKXD-TGS 311
QY 350 PLKVKNSKN-----SKSPEEHLEEMMQWSPNKLHTNFHIPKKGPPAKPKGKSHDKPLK 403
Db 312 P-----SSKNAPLFPSPKWKGEKRIAVGVKKM-----EDAWKRRDLELY 349
QY 404 AKGRS---KGIINGQKSTGNSKSPKGLKTPKTKMKQMTLLDMAKGTQKMTAPRNSGT 460
Db 350 NEACSWCEKNLSQNRST-----FENPIKFSVQKLVDAK----- 385
QY 461 PRTSSKPHKHLPPAALHILAIYKKNOKREDKRSALSCVISTARLLSSDDRARLPEELRS 520
Db 386 -----DRA----- 388
QY 521 LVQKRYELLEHKK--RWASMSSEQRKEYLKKRBEELKKLKEKAKERKEKMLERLEKQK 578
Db 389 -----HMKGMKWAQKAE--FKAWMSSEKRELKYQKFPKIKAWFNEDI----- 428
QY 579 RYEDQELTGKNLPAF--RLVDTPEGLPNTLFGDVAMVVEFLSCYSGLLLPDAQYPITAVS 636
Db 429 ALDDLVTCTENLELLDAKRVNCEDELLKCL-----EISQFFVSMRKILLWNEN--ITAEQ 481
QY 637 LMEALSADKGFLYLN-RVLVILLOTLLQDEIAEDYGEKLGKMLSEIPLTLHSVSELVRLC 695
Db 482 LRDDLHGDFDGFGRSTYKVMIANLLETALQEKHEKAHNCNARLSFFPINEHTISLIRAF 541
QY 696 LRRS-----DVQEESEGS-----TDNDKDSAAAFEDNE 723
```

Db 542 FIGSTSTSKRQKRAARGHDDDDDEEDGDEEMDSREKVVETEEPEENVANGAENGE 601
Qy 724 VODEPLEKLETSBFFELTSEEKLIQILTALCHILMTYSVQDHNMETRQMSAELWKE---- 779
Db 602 IGDDEEENSES-----ELQKQRIALFADSDCHIVELPAGAQ--LEVLICAM-----KEVHVG 650
Qy 780 ----RLAVLKEENDKRAEKQK-----RKEMEAKNKE-----NGKVENGLGKTRDK--- 821
Db 651 LPIREWFWRDANSEKLTTEHKEANRIRNEMEOFHQLOQDFPPIPEITESTMTRTQTRAE 710
Qy 822 ---KRIVKEPEVDTEADMISAVKSRLLAIOAKKEREIQREEMKVKLERQAEERIRK 878
Db 711 QSLRREKLENQDIT-----LKIEENREATAREVD-DLER----- 746
Qy 879 HKAAEAKAFQEGIAKALVMRRTPICGDRNHNRYWLF---SDEVP-----GLFIEKGWVH 930
Db 747 -----IFRVVHIGNDRHLRKYWFAYSSDAIAWVQDFGTTSYEKWYR 788
Qy 931 DSIDYRPNHCKDHTVSGDEDYCPRSKKNALGNASMTQHGTAIVAVETTPKQGNL 990
Db 789 DCSEKGF-----DVESSDVENRPEYEDLP-----TSSQSET 822
Qy 991 WFLCDSOKELDELNLCLHPQIGRESOLKERLEKYODIITHSLARKPNLGLKSCDGNQE 1050
Db 823 WYKLDTEPAIRQLMTOLKNGKREKLLKYLNNMDDIISSILRKEKQ---KDDGEE 879
Qy 1051 -----DEEASEDEASAAENGEEKMETEQNETAAAAAETQAEAKNARRTGRFGSLKRTMSL 939
Db 880 DEEASEDEASAAENGEEKMETEQNETAAAAAETQAEAKNARRTGRFGSLKRTMSL 939
Qy 1063 ATRLQKGGIYVEETSEFPARVISLEKLDKG-----ECVI---ALQASVIKK 1107
Db 940 LNDWKQSGISKIVDSQVFARMLEANTLDEMRLVTELVTSPVECVIEKFPQNVIAKK 999
Qy 1108 FLOGFMAPKQKRKLQSED-----SAKT-----BEVDEEKWVEEAKVASA 1148
Db 1000 FLNFSTILLEKWNVEFSDFRAKCRIFNYFLLFHSKTSKFLKLNQTMKIGAKFGICAI 1059
Qy 1149 LEKWKTA-----IREAQTSRMHVLGLMLDCACIKWDMSAENARC 1187
Db 1060 FQKPKDVRLLPKKKECFSHLKMFRFCRRVQEAENASCLHMLLAYFDARIDQQTLPPELSC 1119
Qy 1188 KYCPKKGEDKUILDECNAKAFHLCPLP-----ALYE-----VPDGEWQCPACOPATARRNS 1240
Db 1120 QVCRKRTGTGERKLMCKQCSTVFHYGCHRPITISRALFEERGFKEG-WWCAKCTKEDRRR-- 1176
Qy 1241 RGRNYTEESASEDSED-----DESDEESEESEEEEDYEVAGLRRLPRKTIRK 1291
Db 1177 -----QLSEAKEDLRQEGGDEEDHGGESDDEEDIVEEETRGRSAKR---- 1224
Qy 1292 HSVIPPAARSGRRPGKPHSTRRSOPKAPP-----VDAEYDELVLQTKRSSRQSLSL 1345
Db 1225 -----KANAAMRDVLVEFGLVRLQTPAPPKPPKQKTVVPEVLELFPNSIERANPRLYKTL 1278
Qy 1346 QKCEILHKIVKYRFSWPFPVTRDEADYDYDVIHPMDFTQVQNKCSGYSRVSQEFL 1405
Db 1279 QQI-----PQCSRSTRNAQHENSRLPDIEQDLV-----YTSAEQLH 1315
Qy 1406 TDMQVFTNA-----EYVYNGRSHVLSVMYK 1431
Db 1316 EHLQOFFRHARGYIETHNPRKFSKFCPPK 1345

RESULT 25

Q23590
ID Q23590 PRELIMINARY; PRT: 1376 AA.
AC Q23590
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Flectin protein 1.
GN Name=flit-1; ORFNames=ZK783.4;

Query Match 5.6%; Score 448.5; DB 2; Length 1376;
Best Local Similarity 18.6%; Pred. No. 5.4e-11; Mismatches 249; Indels 455; Gaps 56;
Matches 286; Conservative 249; Mismatches 550; Indels 455; Gaps 56;

Qy 126 KEKMLKXIVKIHLEKVEDEATEKKSGACDS-----PSSDXENSSQ----- 168
Db 21 QQQLLQQQLAKI---QKATASSPSTKSTNGTSASTSAVSTSTSSSQNEAQLQNLAKWQ 77
Qy 169 -----IAQDHQKKTIV-----KEDEGRRESINDRARRSPRLPTSLKKGERK 211

Db 859 AKLEKER---LEKEAEKRIAEQEKIAEENRIAEQEKIAEEL-----EKKRLQK----- 904

Qy 834 EADDMISAVK-SRRLLATQAKKERIOEREMVKLEROAEE-ERIRKHAAAEAKAFQGI 891

Db 905 EEOURLAAAELEKRLKEAEKRIAEQELEKK-RLEKEAAEVKRIADEAAAAAKLEKE-- 961

Qy 892 AKAKLVMTPTIGTDNRHNRYLWFSDEVPGLFIEKGWVHDSIDYRNFHHCKDHTVSGDED 951

Db 962 -----RLEKEAEKRIADEAAA 977

Qy 952 YCPRSKANLGNAMNTQHGATVAVVETTTTPKQGNLWFLCDSQKEL-----DELLNC 1006

Db 978 AAALKERLEKEAAAEKRIADEAAAEAKLEKE-----RLEKEKRIADEAAAAAAA 1031

Qy 1007 LHPOGI-RESQKERLEKRYODIHSIHLARPNIGLSCDGNQBLNPLRSLDIEVATR 1065

Db 1032 LLOQKIEKEBRDRIAKENKELKEKEDKERQERQERQEKQERARALKKEKIEKER 1091

Qy 1066 LQKGLGVYEETSEFEARVISLEKLKDFGECVIALQASVIKFLQGFVAPKQKRKLOSE 1125

Db 1092 LNQOKLDQEKEREREQERKEQOQRENE-----KQLEKEEREKERREKLKOR 1140

Qy 1126 DSAKTEEVEDEKQVVEEAKVASALEKWK-----TAIREAQTFSRMHVLLGMLDA 1174

Db 1141 NEQLEKERQERFKDOEKEKQLEKQOQOQKVIIPTTTTITIRERS----- 1186

Qy 1175 CIKWMSAENARKVCYKKGEDDKLILDECNKAFHLFCLPALYEVDPGEMQCPACQPA 1234

Db 1187 -----NSDSDALLNLPOASASSSH-----SELVHLTLGRSKAKGRKKP 1224

Qy 1235 TARRNSGRNYTEESASDSE-----DDESDEEEEBEEDYEVAGL----- 1280

Db 1225 TRRELTKDGNRNKNIPNLKELKSTVIPPDDDDQSQOQSEEEVEERFVAVKQNPIGFN 1284

Qy 1281 ---RLRPR---KTIRKGHSVIPPAAASGRPGKPH-----STRRS 1315

Db 1285 PGMMKPDPSAVKLSAKPSTNTTTTAPQOPQVFNLPKVALKPTVTQTTTTTT 1344

Qy 1316 QPKAPPVDDAEV 1327

Db 1345 PPTTTPSSSVQL 1356

RESULT 28

Q8ISF5

ID Q8ISF5 PRELIMINARY; PRT; 10578 AA.

AC Q8ISF5;

DT 01-MAR-2003 (TremBLrel. 23, Created)

DT 01-MAR-2003 (TremBLrel. 23, Last sequence update)

DT 01-MAR-2004 (TremBLrel. 26, Last annotation update)

DE lmda_1 protein.

GN Name=iso1;

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;

OC Rhabditidae; Peloderinae; Caenorhabditis.

OX NCBI_TaxID=6239;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=22269627; PubMed=12311307; DOI=10.1016/S0022-2836(02)00970-1;

RA Flaherty D., Gernert K., Shmeleva N., Tang X., Mercer K.,

RA Borodovsky M., Benian G.;

RT "Titins in Caenorhabditis elegans with Unusual Features: Coiled-coil

RT Domains, Novel Regulation of Kinase Activity and Two New Possible

RT Elastic Regions.";

RL J. Mol. Biol. 323:533-549(2002).

DR EMBL; AV130758; RAN61519.1; --

DR PIR; E89066; E89066.

DR PIR; T33247; T33247.

DR HSP; Q10466; 1BPV.

DR InterPro; IPR008938; ARM.

DR InterPro; IPR003961; FN_III.

DR InterPro; IPR008957; FN_III-like.

DR InterPro; IPR007110; Ig-like.

DR InterPro; IPR003598; Ig_c2

DR InterPro; IPR008941; TFR-like.

DR InterPro; IPR011046; WD40_like.

DR Pfam; PF00041; fn3; 2.

DR Pfam; PF00047; ig; 3.

DR SMART; SM00060; FN3; 2.

DR SMART; SM00408; IGC2; 3.

DR PROSITE; PS50853; FN3; 2.

DR PROSITE; PS50835; IG_LIKE; 11.

SQ SEQUENCE 10578 AA; 1174295 MW; 71BF50C502FB4F16 CRC64;

Query Match 4.6%; Score 363.5; DB 2; Length 10578;

Best Local Similarity 20.2%; Pred. No. 2.3e-06;

Matches 294; Conservative 209; Mismatches 543; Indels 409; Gaps 60;

Qy 25 IPHTQEAFTREEYEARLERYSERIWTCKSTGS---SOLTHKEAWEEQEOVAELKEBFP 81

Db 8976 IGHVQEKTFTAIELVFN-EFYAFRIVAVNGFEGAPSEIIEVNTLDYQE-----ESFD 9028

Qy 82 -AWTEKLVLEWHHTASLEKLVDTAWLEIWKYAVGECDFEVCKEKMVKVIVKIHL 140

Db 9029 FAGEEELKLDVQVNN-----EVVTEITI-ESEVTIEEHKLLKK----- 9068

Qy 141 EKVDEEAETEKKSDGACDSSDKENSSOIAODHOKKETVWKEDEGRRESINDRARRSPRK 200

Db 9069 -----SKLSKTTDEPDLSEIALEVSDITSSLEITTE-----STIPDTAPESQET 9115

Qy 201 LPTSLKGE---RKWAPPKFLPHKYDVKLQNEKIIISNVPADSLIRTPPNKEIVRYFI 257

Db 9116 LNVEIAVTETTVQKITNPSDESADKV---NEDTAVS-----SIVKK----- 9154

Qy 258 RHNAIRAGTGNAPWVEDELVKVSLPSKSFDFLLDPKYMTLNPSTKRKNYTGSPDRKP 317

Db 9155 -----DDKDVNKKSLPES-----GLTTKKEIQGKPEKKI 9183

Qy 318 SKKSTDNSSLSPLNPKLWCHVHLKSLSGSLPKVKNKSNKSPSEEHLEEMKMSPNK 377

Db 9184 MKKTEKADSISETSETLTQDTQKQSEBEPAK---RTTETSVDQEVKRTETTSKSK 9240

Qy 378 LHTNFHPKPGPPAKPGKSHDKPLKAKGRSGILNGQKSTGNSKSPKGLTKPTKMKQ 437

Db 9241 QTTEEH-----PQPGKSDSSISSTDASEVKQVQOSEAQKVPETAKLESK- 9291

Qy 438 MTLDMAGTKOTMTR-APRNSGGTPTSSKPHKHLPPAALHLIAYYKKNKOREDSALS 496

Db 9292 -----SKMTEDTTKESDNKETVDEKPKKYV-----LKKKTEKSDS----- 9326

Qy 497 CVISKTRLLSSEDRARLPEELRSIVQKRYELLEHHKRWASMSERQR--KEYLKKREEL 554

Db 9327 -TITSETSVAVESAGPSESTQNV-----AVDKEKKQKETDEKQKLEAEIAGKSTEQ 9380

Qy 555 KKKLKEKAKERREKEMLERLEKQRYEDQETGKNLPAPRL-VDTPEGLPNTLFGDVAVV 613

Db 9381 KSKLEAEAKKRAAEE-DAAKKQK--EKTEAASKAAAEKLEKLEKQAOINKAAAEADAVKK 9437

Qy 614 VEPFLSCYGLLLPDAQYFITAIVSL-MEALSADKGGFLYNRVLIVLLQTLQDETAEDYG 672

Db 9438 QNELDEQNKL---EATKKLAAEKLEKLEQSAK-----SKQAAEBQA 9476

Qy 673 ELGKMLSEIPLTLHSVSELVLCRLRRSDVQSESGSDTDD--NKDSAAAFEDNEVODE--- 727

Db 9477 KLDQAQ-----TKAAAEKQTLGKDEKSNKDSGS---NETVEEKP 9514

Qy 728 -----FLEKLETSEPFELTSEKQLIALTALCHRLIMTYSVQDHMETRQOMS 773

Db 9515 KVKLKKTEKSDSSISQSDTSK-----TVAESAGSSESTQKVADATSKQKETDKQKLE 9570

Qy 774 AELWKEKRLAVLKEENDK-----KRAEKQKREMEAKNKNKENGKVENGLGTDKRLIV 825

Db 9571 ABITAKKSA---DEKSKLETESKLKAAEADAQKQEKEDKLKLEADVASKAAAEKLEL 9627

Qy 826 KFEPOVDTEADMISAVKSRRLAIQAKKREIQEREM---KVKLERQ-----AEEERI 876

Db 3705 AIDEKSQAENVSEIVSEKITDEKAQESQKEVKDSEAKPKKAKVLEKKSIEBEKLEDKKE 3764
Qy 154 GACDSPSSDKENSSIOAQDHQKKEVTVKEDGRRSINDRARRSRPKLPTSLKKG--RRK 211
Db 3765 TOTDSAIIDEKSQAENVSEIVSEKITDEKAQESQKEEVKD--SEAKPKKAKVLEKKSIEBEK 3823
Qy 212 WAPPKFLPHKVDVQLQNDKIIISNPADSLI--RTERPPNKEIVRYFRHNALRAGTCEN 269
Db 3824 LEDKKEKOTESAIDEKSQAENVSEIVSEKITDEKAQESQKEV-----KGSE 3870
Qy 270 APWVVEDELVKYISLPSPFLLDPYKYMTLNPSTKRKNTGSPDRKPSKSK----- 322
Db 3871 AKPKKAKVLEKKSIEBEKLED-----KKEKOTESAIDEKSQAENVSEIVSE 3916
Qy 323 --TDNSSLSPL-----NPKLWCHVHLKKSLSGSPKVNKSNKSPSEHLEEMKM 372
Db 3917 KITDEKAQESQMEVEVKDSEAKPKK--AKVLEKKSIEBEKLE---NKKEKOTESAIDEKSQK 3972
Qy 373 MSPNKLHNFHPIPKKGPAP--KPGKHSD--KPLKAGRSKGILNGOKSTGNSKSPKGLKT 430
Db 3973 AEVSEIVSEKITDEKAQESQKEVKDSEAKPKKAKVLEKKSIEBEK-----LED 4021
Qy 431 PKTKMKQMTLMDMAKGTQMTRAPNSGGTPTRTSKPHKHLPPAALHLIAYYKKNKRED 490
Db 4022 KKEKOTESAIDEKSQAENVSEIVSENI--TDEKAQESQK-----KEVKDSEA 4066
Qy 491 KRSALSCVISTARLLSSDRARLPEELRSIVOKRYE-----LLEHKRWASMSG-- 540
Db 4067 K-----PKKAKV-----EKKSIEBEKLEDKKEKTESAIDEKSQAENVSEI 4108
Qy 541 -----BORKEYLK-----KKEELKKKL--KEKAKERREKEMLERL--EKOKRY 580
Db 4109 VSEKITDEKAQESQKEEVKDESEAKPKKAKVLEKKSIEBEKLEDKKEKOTESAIDEKSQA 4168
Qy 581 EDQELTGKLPAPRLVDTPEGLPNTLFGDVAMVVEFLSCYSGLLPDQAQYPIITAVSLMEA 640
Db 4169 EVSEIVSENI-----TDEKAQESQKEV-----KDESEAKPKKAKVLEK 4206
Qy 641 LSADKGGFLYLNVLVILLQTLQDEIADYGEIGMKLSEIPLTHS-----VSELVRLCL 696
Db 4207 KS-----IEBEKLEDKKE--KOTESAIDEKSQAENVSEIVSEKI 4243
Qy 697 RRSVDQESGSDTDDNK-----DAAAFEDNEVODE-----FLEKLETSEFFE 739
Db 4244 TEEKQAQ--ESQKEVKDSEAKPKKAKVLEKKSIEBEKLEDKKEKTESAIDEKSQAENVSE 4302
Qy 740 LTSEKQLIALTALCHRIILMTYSVQDPMETROQMSAELWKERLAVLKEENDKRAEKQKRK 799
Db 4303 IVSEK-----ITDEKAQESQ-----KEEVKDSEAKPKKAK 4332
Qy 800 EMEAKNKGKVENGLGKTRDKRIRIVKPEPOVDTEADM--ISAVKSRLL--AIQAKKER 856
Db 4333 VLEKKSIEBEKLEN-----KKEKOTESAIDEKSQAENVSEIVSEKITDEKAQESQK 4384
Qy 857 BIOEREMVKV--LEROA--BEERIR--KHAAAEEKAFQEGIAKAKLVMMRPTIGTDRHN 910
Db 4395 EVKSEAKPKKAKVLEKKSIEBEKLEDKKEKOTESAIDEKSQA----- 4428
Qy 911 RYWLFSDEVPLGFIETKGWVHDSIDYRFNHHCKHDTVSGDEDYCPRSKKNALGKNASM--- 967
Db 4429 -----EVSEIVSEK-----ITDEKAQESQKEV-----KDESEAKPKKAKVLEKKSIEBEA 4472
Qy 968 -----NTQHTGATEVAVETTTTQGGQNLWFLCDSQKE-----LDELL 1004
Db 4473 KLEDKKEKTESAIDEKSQAENVSEIVSEKITDEKAQ-----ESQKEEVKDSEAKPKKA 4526
Qy 1005 NCLPHQGRRESQKLERLEKRYQDIIHSTHARKNIGLKSQDGNQELNLFRLSLIEVAT 1064
Db 4527 KVLEKKSIEBEKLEDKKEKTESAID--EKSQKAEV-----SEIVSEKITD--EKAQ 4574
Qy 1065 FLQKGGGLGYVEETSEFEA-----RVISLEKLDKFGECVIALQASVIKFLQGFMA 1114
Db 4575 ESQK-----BEVKDSEAKPKKAKVLEKKSIEBEKLED-----KKEKOTESA 4615

Qy 1115 PKQKRKLQ-----SEDSAKTEEVDEKKNVVEAKVASALEKMKTAIREAQTFSRM 1165
Db 4616 IDEKSQAENVSEIVSEKITDEKAQESQMEVEVKDSEAKPKKAKVLE--KKSIEEAKLEDKK 4673
Qy 1166 HVLGLMDACIKWDSMAENAR--CKVCPKKGDDKLILCDECNKAFHFLCLRPALYVEPDG 1224
Db 4674 ET--QTDSAI--DEKSQAENVSEIVSEKITDEK--AQESQKE-----EVKDS 4714
Qy 1225 EWQCPACOPATA-----RRNSGRNYTTESASEDEDESDB--EEEEEEEEE 1270
Db 4715 E-----AKPKKAKVLEKKSIEBEKLEDKKEKOTESAIDEKSQAENVSEIVSEKITDEKAQ 4769
Qy 1271 EEDYEVYAGLRLPRKT--IRGKHSVIPPAARSGRRPGKPKPHSTRRSQKAPVDD----A 1325
Db 4770 ESQKKEVKGSEAKPKKAKVLEKKS-----EKEKLEDKKEKOTESA-----IDEKSQA 4818
Qy 1326 EVDSLVLQ-----TKRSSRRQSLELQKCEE-----ILH 1353
Db 4819 EVSEIVSEKITDEKAQESQKEVKDSEAKPKKAKVLEKKSIEBEKLENKKEKOTESAIDE 4878
Qy 1354 KIVKYRSWPPREPVTREADEYDVTHTPMDFTQVQNKCSGYSRSV--QBEFLTDMK--- 1409
Db 4879 KSQAENVSEIVSEKITDEKAQESQK--KEVKDSEAKPKKAKVLEKKSIEBEKLEDKKEKQ 4936
Qy 1410 -----QVFTNAEV-----YNCRGSHVLSCHVMTQCLVVLHKLHPGHPYV 1450
Db 4937 TESAIDEKFKQAEVSETVSEKITDEKAESRKEEVKDESEAKPKKAKV----- 4984
Qy 1451 RRKKKPFDRLAEDGSEPEAVGQSDRDERRSREAEIQEWLQD 1494
Db 4985 --EKKSIEBEKLEDKKEKOTESA-----IDEKSQAENVSEIVSE 5021

RESULT 33
Q80U42 PRELIMINARY; PRT; 886 AA.
AC Q80U42, 2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE MKTAA0314 protein (Fragment).
GN Name=Baz2a; Synonym=mkTAA0314;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=22579291; PubMed=12693553;
RA Okazaki N., Kikuno R., Ohara R., Inamoto S., Aizawa H., Yuasa S.,
RA Nakajima D., Nagase T., Ohara O., Koga H.;
RT "Prediction of the coding sequences of mouse homologues of KIAA gene:
RT II. The complete nucleotide sequences of 400 mouse KIAA-homologous
RT cDNAs identified by screening of terminal sequences of cDNA clones
RT randomly sampled from size-fractionated libraries.";
RL DNA Rea. 10:35-48(2003).
DR EMBL; AK122243; BAC65525.1; -.
DR HSSP; Q9UG10; 1F62.
DR MGD; MGI:2151152; Baz2a.
DR GO; GO:0006338; P:chromatin remodeling; TAS.
DR InterPro; IPR000637; A:T_hook.
DR InterPro; IPR001487; Bromodomain.
DR InterPro; IPR011011; FYVE PHD_Znf.
DR InterPro; IPR001965; Znf_PHD.
DR Pfam; PF02178; AT_hook; 2.
DR Pfam; PF00439; Bromodomain; 1.
DR Pfam; PF00628; PHD; 1.
DR SMART; SM00503; BROMODOMAIN.
DR SMART; SM00384; AT_hook; 2.
DR SMART; SM00297; BROMO; 1.
DR SMART; SM00249; PHD; 1.


```

DR PROSITE; PS50014; BROMODOMAIN 2; 1.
DR PROSITE; PS50016; ZF_PHD_2; 1.
FR NON_TER 1
SQ SEQUENCE 886 AA; 100060 MW; 43E1172F5F081E06 CRC64;

Query Match 4.4%; Score 353; DB 2; Length 886;
Best Local Similarity 21.8%; Pred. No. 3.8e-07;
Matches 114; Conservative 78; Mismatches 188; Indels 144; Gaps 14;

QY 975 TEVAVETTTKQGNLWFLCDSQKELDELINCLHPQCIRESQLEKRYQDIHSHL 1034
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
408 TQLTAQPIPEMCSGWWIRDP- TLDVLLKALHPRGIREKALHKLKSK-HKDFLEVC 465
QY 1035 ARKPNLGLKSCDGNQELLN--FURSLIEVATLQKGLGYVETSEFARVLSLEKLD 1092
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
466 : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1093 FGECVIALQASVIKKFLOGMAP----- 1115
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
508 LEQVVLSLQI-----RGWTCPTDPSTREDLTYCHLPDSPEDIPWRCRGREGTVPQRQ 562
QY 1116 -----KQKRKLQSDSAKTEVDEBKWVE-----BAKVA 1146
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
563 NNPLDLAVRLAVLQNVERRYLREPLWAHAEVVVEKALLSTPNCAPDGTSTEISYEIT 622
QY 1147 SALEKWKTAIREAQTFSRMHVLLGMDACIKWDSAEANARCKVCPPKGGEDDKLILCDECN 1206
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
623 PRVRVWRQTLERCRSAQAQVCLMGQLERSIAWEKSVNKVTCVLCVRKGDNDDEFLLCDGCD 682
QY 1207 KAPHLCLRALPYEDPGWQCPACOPATARNRSGRYTEESASEDSDESDDEEBEE 1266
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
683 RGCHYICHRPMKAVPEGDMFCVCL----- 708
QY 1267 EEEEDYEVAGLRRLPRKTIKGHSVIP---PAARSGRRPGCKPHSTERSOPKAP--P 1321
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
709 -SQVEEYQRPFPKQK-----RKSSPLTFPEGDSRRR-----MLSRSRSPAVPRVP 759
QY 1322 VDAEYDELVLQTKRSRRQSLELQKEILHKIVKRFSPWPFPVTRDEAEYDVIT 1381
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
760 EDGLSPK---RRHSMRSHSLTFCETILMEMESHDAAPLEFPVNPRLVSGYRRVVK 816
QY 1382 HPMDFTQVQKSCGYSRVSQBELTDMQVFTNAEVNCRGSHV 1425
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
817 NPMDFETMRERLLRGYTSSEFAADALLVDFNCQTFNEDDSEV 860

RESULT 34
ITCL_YEAST STANDARD; PRT; 1264 AA.
AC P53125;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Imitation switch two complex protein 1.
GN Names=ITCL; OrderedLocusNames=YGI133W; ORFNames=G2842;
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / FY1679;
RX MEDLINE=96437978; PubMed=8840506;
RX DOI=10.1002/(SICI)1097-0061(199607)12:9<887::AID-YEA971>3.3.CO;2-4;
RA Escribano V., Eraso P., Portillo F., Mazon M.J.;
RT "Sequence analysis of a 14.6 kb DNA fragment of Saccharomyces
RT cerevisiae chromosome VII reveals SEC27, SSM1b, a putative S-
RT adenosylmethionine-dependent enzyme and six new open reading frames.";
RL Yeast 12:887-892(1996).
RN [2]
RP FUNCTION, INTERACTION WITH ISW2, AND SUBCELLULAR LOCATION.
RX MEDLINE=1137229; PubMed=11238944;
RX DOI=10.1128/MCB.21.6.2098-2106.2001;

```

```

RA Gelbart M.E., Rechsteiner T., Richmond T.J., Tsukiyama T.;
RT "Interactions of Isw2 chromatin remodeling complex with nucleosomal
RT arrays: analyses using recombinant yeast histones and immobilized
RT templates.";
RL Mol. Cell. Biol. 21:2098-2106(2001).
RN [3]
RP FUNCTION.
RX MEDLINE=22511509; PubMed=12624196; DOI=10.1099/mic.0.25920-0;
RA Ruiz C., Escribano V., Morgado E., Molina M., Mazon M.J.;
RT "Cell-type-dependent repression of yeast a-specific genes requires
RT Itclp, a subunit of the Isw2p-Itclp chromatin remodeling complex.";
RL Microbiology 149:341-351(2003).
CC -!- FUNCTION: Required for nucleosome-stimulated ATPase activity and
CC chromatin remodeling activity of the ISW2/ITC1 complex. Required
CC for the repression of MATa a-specific genes.
CC -!- SUBUNIT: Component of the ISW2/ITC1 chromatin remodeling complex.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: Contains 1 DDT domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL; Z72655; CAA96844.1; --
DR PIR; S64146; S64146.
DR Germonline; 141181; --
DR SGD; S00003101; ITC1.
DR GO; GO:0008623; C:chromatin accessibility complex; IPI.
DR GO; GO:0005634; C:nucleus; IDA.
DR GO; GO:0006348; P:chromatin silencing at telomere; IMP.
DR GO; GO:0046020; P:negative regulation of transcription from P. .; IMP.
DR InterPro; IPR004022; DDT.
DR InterPro; IPR009058; Wing_hlx_dna_bnd.
DR Pfam; PF02791; DDT; 1.
DR SMART; SM00571; DDT; 1.
DR PROSITE; PS50827; DDT; 1.
DR Chromatin regulator; Nuclear protein; Repressor;
KW Transcription regulation.
FT DOMAIN 423 483 DDT.
SQ SEQUENCE 1264 AA; 145642 MW; 45E4CF8835C7C746 CRC64;

Query Match 4.4%; Score 347.5; DB 1; Length 1264;
Best Local Similarity 19.5%; Pred. No. 9.8e-07;
Matches 271; Conservative 206; Mismatches 492; Indels 419; Gaps 58;

QY 3 PLLGRKPPPLVNPPLGPEEPFFTIPTQAFRTREVEARLERYSERIWTCKSTGSSQLTH 62
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
8 PILLPDPKPL--PLDLNVQVWHIEETGEWFSSEYEFLEFRFDYTRHHFTCEITGTSCLTF 65
QY 63 KEAWEEBEQVLAELKEEFPAPWYEKLVLEMVHN-TASLEKLVDTAWLEIMTKYAVGEED 121
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
66 FOALDSEETQFKYVEDRFPFLKRLREPVARPLHFNGIRRLDALVEKVFARFNDFFPGE--- 122
QY 122 FVPGKEKMLKVKVILHPLEKVDTEATEKSGDAGCDSSDDKENS-----QIAQDHQKE 177
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
123 -----VYLRKQKSDSTTSNSQSTFPQDDMWVEINSVGNPGLPQYQYQRR 168
QY 178 TVVKEDEGRRESINDRARRSRPKLPTSLKGRKRWAPPKFLPKHYDVKLQNDKTIISN-- 235
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
169 YVYKEKQVFNATINESIEIVMPAHT-----KY---MLIEAASSNKS 208
QY 236 --VPADSLIRTPPNKKEIVRYFIHNNALRAGTGNAPWVBEDELVKYKSLPSKFSDFLL 293
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
209 FIVDQGIYVDRDSTFTKHLIKCFKXITLQIRASSKMGAPWCVPKPEYLAWYGLTMEWPK--- 265
QY 294 DPYKVTMLNPSTYRK-----NTGSPD-----RKPSKSKSTDNSSSLNPLKLVCHVL 342
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : :
266 DMLKYKEDEFPVVARNSANVSSPESKKNKRSKSGKSNSTNSDASNKETKK----- 318

```


DE CG10897-PA.
 CN Name=tou; ORFNames=CG10897;
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RN SEQUENCE FROM N.A.
 RX MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.H., Blaise R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
 RA Abell J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballwe R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kianos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.Y., Wasmann D.A., Weinstein G.M., Weissenbach J.,
 RA Williams S.M., Woodgett, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
 RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*.";
 RN Science 287:2185-2195(2000).
 RN [2]
 RN SEQUENCE FROM N.A.
 RX MEDLINE=22426065; PubMed=12537568;
 RA Celniker S.E., Wheeler D.A., Kronmiller B., Carlson J.W., Halpern A.,
 RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,
 RA George R.A., Hoskins R.A., Laverty T., Muzny D.M., Nelson C.R.,
 RA Pacleb J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,
 RA Svirskas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
 RA Weinstein G., Scher S.E., Myers E.W., Gibbs R.A., Rubin G.M.;
 RT "Finishing a whole-genome shotgun: Release 3 of the *Drosophila*
 RT melanogaster euchromatic genome sequence.";
 RN Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).
 RN [3]
 RN SEQUENCE FROM N.A.
 RX MEDLINE=22426070; PubMed=12537573;
 RA Kaminker J.S., Bergman C.M., Kronmiller B., Carlson J., Svirskas R.,
 RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,
 RA Ashburner M., Celniker S.E.;
 RT "The transposable elements of the *Drosophila melanogaster* euchromatin:
 RT a genomic perspective.";
 RN Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).
 RN [4]
 RN SEQUENCE FROM N.A.
 RX MEDLINE=22426069; PubMed=12537572;
 RA Mista S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,

RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,
 RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
 RA Bettencourt B.R., Celniker S.E., de Grey A.D., Drysdale R.A.,
 RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
 RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
 RA Lewis S.E.;
 RT "Annotation of the *Drosophila melanogaster* euchromatic genome: a
 RT systematic review.";
 RN Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).
 RN [5]
 RN SEQUENCE FROM N.A.
 RP SEQUENCE FROM N.A.
 RG FlyBase;
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
 RN [6]
 RN SEQUENCE FROM N.A.
 RP FlyBase;
 RG Submittted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 RL EMBL; AE003825; AAF58638.2; -.
 DR HSSP; O9UIG0; 1P62.
 DR FlyBase; FBGN0033636; tou.
 DR GO; GO:0003677; F:DNA binding; IEA.
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
 DR InterPro; IPR000194; ATPase a/bcentre.
 DR InterPro; IPR001487; Bromodomain.
 DR InterPro; IPR004022; DDT.
 DR InterPro; IPR001739; Methyl-CpG_bind.
 DR InterPro; IPR006162; Ppantne S.
 DR InterPro; IPR001965; Znf_PHD.
 DR InterPro; IPR001876; Znf_RangDP.
 DR Pfam; PF00439; Bromodomain; 1.
 DR Pfam; PF02791; DDT; 1.
 DR Pfam; PF01429; MED; 1.
 DR Pfam; PF00628; PHD; 2.
 DR PRINTS; SM00503; BROMODOMAIN.
 DR SMART; SM00297; Bromo; 1.
 DR SMART; SM00249; PHD; 2.
 DR SMART; SM00547; Znf_RBZ; 1.
 DR PROSITE; PS00152; ATPASE ALPHA_BETA; UNKNOWN_1.
 DR PROSITE; PS50827; DDT; 1.
 DR PROSITE; PS00012; PHOSPHOPANTHEINE; UNKNOWN_1.
 DR PROSITE; PS01359; ZF_PHD 1; UNKNOWN_1.
 DR PROSITE; PS50016; ZF_PHD 2; 2.
 SQ SEQUENCE 3080 AA; 330285 MW; 32A4D606F071AB3 CRC64;
 Query Match 4.3%; Score 343; DB 2; Length 3080;
 Best Local Similarity 15.6%; Pred. No. 4.2e-06;
 Matches 345; Conservative 305; Mismatches 532; Indels 1036; Gaps 82;
 QY 69 EQEVAELLKEFPAYWEKLV-----LEMVHNTASLEKLVDTAWLEIM 111
 DB 1003 EQQSNLSRENFSSARAIVGSFLOPPPPYANDGEYIRMTDEVA--KRLED---LKVF 1057
 QY 112 TKYAVGECDCEVGVKEK--MLKVKIVKTHPLEKDEEATEKKSGACDPSDDKENSQI 169
 DB 1058 TQTLNVQRIEIAQQQAMEDAKVLOKEELARNKEKARQEK-NSKLQQRKDKELKNQ 1116
 QY 170 AQDHQKKE-----TVVKEDEGRRESINDRARRSPKRLPTSLKKGERKWAAPPFLPHKYD-- 223
 DB 1117 AVEERERRRRQHMSLRMLELRK-PEDREKKKHQIVLDRLRLRRRMA-----ERKRD 1170
 QY 224 ----VKLONEDKIISN---VPADSLIRTERPNKE-----IVRYEIR----- 258
 DB 1171 ILQLIRRNEDSEMPQELVIPELORIAGNRUPGQAMDLLMVFELHFNGETLGFDMESL 1230
 QY 259 -----HNALRAGTGENAPWVVEDEL-----VKYKSLP----- 285
 DB 1231 PSLQNLHDLMSDSNADA----EEELLSVMTLLVCAIEDPCVPNPGRHHTLLGQSLRNA 1286
 QY 286 ----SKFSDFLDPYKYMTLNPSTKRNKTGSPDRKPSK----- 320
 DB 1287 DITNSNVSE-ILRIYLYATATGCVQRMHGITVDRERERRVDPDHOLDSDTTTTHSHSVKNQ 1345

QY 321 -----SKTNSLSS-----PLNP-----KLMCHV----- 340
Db 1346 EYKLLHENDTWKLSQSKDRPFVALNPTKQMLAHLCDLLNKNKAVLQIDGSLTCA 1405
QY 341 HLKSLSGSPKVKNSK-----NSKSPHEELENMKNKSPNKLH----- 379
Db 1406 QMRKEKMTDMKVKYKALHMRKARIEAYERAQAREAMQALMAQOKLDAERLKABEEA 1465
QY 380 -----TNFHIPKGGPP-----AKPGKHSKPLKAKG----- 406
Db 1466 KAAAAEAAAAAGTDCGATKGGSPNGEKPEDGQNEEGRAKEPQOQOQPMVEVDGVDEA 1525
QY 407 -----RSKGLNGOKSTGNSPKK-----GLKTPKTKKQMTLLDMAKGTQKMT 451
Db 1526 SLVSPAKTII-----QTDNLSLTPSKODMPTPTQINGSSSTPTTS-----GVTGDMNVL 1574
QY 452 RAPNSGGTPTSCKPHKLPAPALHL-----IAYKENKORE-----DK--RS 493
Db 1575 LQAKKSGARNSINDEHHHVSIIIDDLSDLSDEITNVEEDEDNRLSADELQKLDKIVRA 1634
QY 494 ALSC--VISKTARLLSS-----EDRA-----RLPEE-----LRSLVQKRYELLEHKRWASW 538
Db 1635 SLNCKEALSKSTQLRAACFCQDRFWRRYWKLPKAGGIFIEALESQNDICDYHEALEAM 1694
QY 539 SEQRKEYLKKRBEELKKLKEKAKERRE--KEMLERLE-----KQKRYEDQ 593
Db 1695 DDKDANDEKENSENEKDVAAESSEQPMVEVDESIKLEDGVPASDVGMPPESNQNAHQDE 1754
QY 584 ELTKNLPAPRLV-----DTPGCLPNTLFGDVAMVVEFLSCYSGLLLPDAQVP--- 631
Db 1755 EDDDDVTEINKVEPBIVDLGDODDDAAPP-----LPIEPPE 1794
QY 632 ITAVSLMEALSADKGGFLYLNRLVILLQTLLODEIAEDYGLGMLSLBIPLTHSVSEL 691
Db 1795 IKVSEMLMGPPP-----TTMISTKTDFAEI-----KIPSPGILMPTT-- 1835
QY 692 VRLCLRRSDVQESSEGSDDDNKDSAA-----FEDNEVQDFLEKL----- 732
Db 1836 --LNNNTNNNNNGSNDCKLETGLGLQOQOQNFSSQVKTEDVKKEDDCIIVSTSSV 1893
QY 733 -ETSSEPP-----ELTSEE-----KLOILTA--LCHRLMTYSVDH----- 765
Db 1894 DDTFKWFSIVRREVPLISELPAEBSGGVGOELQISYANQNCSAQL---QLQGHMPDLINN 1950
QY 766 -----METROQMSAELWKE---RLAVLKEENDKKRAEKKRKEAKNKENGKGVENGL 815
Db 1951 MQYYSIPMBECKVDTSKLGNECIFSLSGDEKQMLAKVBEYKAHVESK-----NGL 2002
QY 816 G-----KTRDKRIVRPEPOVDTE---ABDMISAVKSRLLA--IQAKKERIEQ--R 861
Db 2003 GSPHRHETKDDBEQAKLKDKEIDTEMETDADDLAGKEFFELRSDVPPDTGGVSEGT 2062
QY 862 EMVKLERQAEERIRKHAABKAFQEGIAKAKLVNRRIT-----PIGTDRNHNRYWL- 914
Db 2063 DVKPKTELRLDE-----ALSOAYYHNIANMSLSSVQTYPIPIDPLSLMTPDHRLL 2114
QY 915 -----FSDVPGFLIEKGWVHDSIDYRFNHHKCHDHTVSGDESDYCPRSKANLGNASW 967
Db 2115 EQVKLAGFPERVHGVTV----- 2131
QY 968 NTOHGTATEVAVETTTPKQONLWFLCDQKELDELLNCLHPQIGRESQLEKRYQD 1027
Db 2132 -----PRQRYGWQLDDEQKLRLQTLNPSGLERELQENLQR--- 2171
QY 1028 IHSIHLAKPNLGLKSCDGNQBELNF--LRSDL-----IEVATRLQKGG 1070
Db 2172 -----FLGLEQPLG---VNYKLKSDIDFPPEFLMPDKGMNPKVAKRVE--- 2213
QY 1071 LGVVEETSFEARV-----ISLEKLKD-----FGECVTA 1099
Db 2214 LALIEQLESLEDAVASAQMLQWQLPNRVESELTLDSQEDVTEDFVSIIPMIRERIID 2273
QY 1100 LQASVKKFLOQFMAPK-----QKRRKLQSEDSAK 1129

Db 2274 LEANIERRYKPLPGSQTDGAHLAVIAQNOHTTTTQNSASAAAYLLOQOQOQOQLAQ 2333
QY 1130 TEEVDE-----EKQVEEAKVASA----- 1148
Db 2334 QOQOQOQOQAGNSLNPSPFNERTWALAAAAAASPGNATGVANSVAVAGATPCESGSGE 2393
QY 1149 -----LEKWKTAIREAQTFSRMHVLLGLMDACIKWDMSS- 1181
Db 2394 PNSGNASPASNCSDRDERKEQI PKGLVQWRDAVSRSHHTAQLAMALYVLESVAVDKSI 2453
QY 1182 ----- 1181
Db 2454 MKAYATKNSKSKKSAKQATPSKKQOQKNGKEQQLTPNGKESKSAINKPENQAPL 2513
QY 1182 -----AENAR----- 1186
Db 2514 SIKINLKALAQNGQCLLTKYTPPIUTKSPTASPSHPNTSDSDSGFKTKKXGSGKRRR 2573
QY 1187 -----CKVCPKKGGDDKLIILCECNKAFHLFCRLPALYEVDPGEWQCCPA 1230
Db 2574 SANNTNSSKYSNSLQNCQFCSTGENEDKLLLCGCDKGHYTCFKPKMDNI PDGDWYCYE 2633
QY 1231 C-OPATARN-----SRGRNY----- 1245
Db 2634 CVNKATNERKCI VCGHRPSPVGMKYCDLCPRAVHADCYIPPLKVPGRKWKYCHGICSR 2693
QY 1246 -----TEESASEDSEDE-----SDEEEEEEEEEEEEDYEVAG----- 1279
Db 2694 APPPKRSAGTSGSSSKSRDRDRDRESGSAKRSRDSNKTTPAMEHMQOQOQPLAGDSHH 2753
QY 1280 -----LRLRPKRTIRGHSV-----IPP 1297
Db 2754 HHQOQPPSLNSHSDSMNSLPAAPLSPAHSVVVSATNYDDOHHANNSVDGSSRFHAHLIP 2813
QY 1298 A-----ARSGR- 1303
Db 2814 SNGTAALLDVEGGANVMGPVYVPTVVAAGNFSAGLINQAPQAMPFANVAVMSRA 2873
QY 1304 -----RPGKKPHS----- 1311
Db 2874 VTPTRTRTPPTPAPTPPPPTPLMQASPTATALHVNACOSPQOQHAQLMTWSPPA 2933
QY 1312 -----TRRSOKAPPVD---DAEVDLVLQTKR---SSRQSLELOKCEBILHKIVKRYFS 1361
Db 2934 IGVGTATQMSPPPINIHAIQAEKLEKEKEKATKMLKELAVCKTLGEMELHEDS 2993
QY 1362 WPFEPVTRDEADYDVIDITHPMDPQTQVONKCSGYSRVSQBFITDMKQVFTNAEYVN 1419
Db 2994 WPFLLPVNTKQPPTYRKIIKTMDLSTIKKQLQDLSYKTREDFCVDVRQIFDNCMFEN 3051
RESULT 37
QY5PR8 PRELIMINARY; PRT; 1637 AA.
AC Q75PR8;
DT 05-JUL-2004 (TReMBLrel. 27, Created)
DT 05-JUL-2004 (TReMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TReMBLrel. 27, Last annotation update)
DE Unichrom.
OS Hemictrotus pulcherimus (Sea urchin).
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC Echinoidea; Euechinoidea; Echinacea; Echinoida; Strongylocentrotidae;
OC Hemictrotus.
OX NCBI_TaxID=7650;
RN (1)
RP SEQUENCE FROM N.A.
RA Moritani K., Tagahira H., Shimotori T., Tanaka S., Takata K.,
RA Mitsunaga-Nakatsubo K., Shimada H., Sakamoto N., Yamamoto T.,
RA Akasaka K.,
RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB164414; BAD12142.1; -.
GO; GO:0003677; F:DNA binding; IEA.

DR HSP; Q9UIG0; 1F62.
 DR FlyBase; FBgn0033636; tou.
 DR GO; GO:0003677; F:DNA binding; IEA.
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
 DR InterPro; IPR000194; ATPase a/bcentre.
 DR InterPro; IPR001487; Bromodomain.
 DR InterPro; IPR004022; DDT.
 DR InterPro; IPR001739; Methyl-CpG_bind.
 DR InterPro; IPR006162; Plantane S.
 DR InterPro; IPR001965; Znf_PHD.
 DR InterPro; IPR001876; Znf_RangDP.
 DR Pfam; PF00439; Bromodomain; 1.
 DR Pfam; PF02791; DDT; 1.
 DR Pfam; PF01429; MBD; 1.
 DR Pfam; PF00628; PHD; 2.
 DR PRINTS; PR00503; BROMODOMAIN.
 DR SMART; SM00297; BROMO; 1.
 DR SMART; SM00249; PHD; 2.
 DR SMART; SM00347; Znf_RBZ; 1.
 DR PROSITE; PS00152; ATPASE ALPHA BETA; UNKNOWN_1.
 DR PROSITE; PS50014; BROMODOMAIN_2; 1.
 DR PROSITE; PS50827; DDT; 1.
 DR PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN_1.
 DR PROSITE; PS01359; ZF_PHD_1; UNKNOWN_1.
 DR PROSITE; PS50016; ZF_PHD_2; 2.
 DR SSEQUENCE 3109 AA; 333242 MW; DDOA75215541DCED CRC64;
 SQ

Query Match 4.2%; Score 332.5; DB 2; Length 3109;
 Best Local Similarity 15.5%; Pred. No. 1.2e-05; Gaps 83;
 Matches 348; Conservative 305; Mismatches 529; Indels 1065; Gaps 83;

QY 69 EQVAVELKEEFPAWYEKLV-----LEMVHNTASLEKLVDTAMLEIM 111
 DB 1003 EQQPSLNRENFSAFVGSFLQAPPYANDGEVIRMTDEVA--KELED---LKVF 1057

QY 112 TKYAVEECDFVGKX--MLKVIKIHLEKVDSEATEKSGDAGCDSPSSDKENSOI 169
 DB 1058 TROTLANVEIRIETAKQOAMRDACKQKELARNEKARQEK--NSKLEQQRKDKELKNOQ 1116

QY 170 AQDHQKE---TVVKDEGRRESINDRAARRSPKLPSTLSKGERKAWPKPLPHKYD-- 223
 DB 1117 AVEERRRRQHMSLRMLELRRK--FEDREKKHQVLVLDLLRERMA-----ERKRDAB 1170

QY 224 ----VKLQNEDKLIIS--VPADSLIRTERPPNKE-----IVRYFIR----- 258
 DB 1171 ILQLIRRPNEDEMPQELVPELDRIAGNLPQAMADLLMVFELHNFGETLGFPMESL 1230

QY 259 -----HNALRAGTGENAPMWVEDEL-----VKYSLP----- 285
 DB 1231 PSLQNLHDALMSDSNADA---EEELLSVMTHLLVCAIEDPGVPNPGRHTLLGQSLRNA 1286

QY 286 ----SKFSDLLDPKYMTLNPSTKRKNTGSPDRKPSK----- 320
 DB 1287 DITNSNVSE--ILRIYLYATATGEVRQMHGITVDRERRRVPDHHQDSDTTTHSHSVKQ 1345

QY 321 ----SKTDNSLSLSS-----PLNP-----KLWCHV----- 340
 DB 1346 EYYKLLHENDYKLSQSLKDRPPVALNPTRKAQMLAHLNDLLMKNKAVLRQIDGSLTCA 1405

QY 341 HLKKSLSGSPKVKYNSK-----NSKSPPEHLEEMKMKMSPNKLH----- 379
 DB 1406 QMRKEKYMTDMKVRKYKALHMRKARIEAVERAQAEREAQAALMAQQLDAERLKAEEBA 1465

QY 380 -----TNFHIKKGPP-----AKPKGSIDKPLKAG----- 406
 DB 1466 KAAAEAEAAAAGTGEATKGGSPNGEKPEDGDQNEEGAAKEPQQQQQPFMEYDGVVDEA 1525

QY 407 ----RSKGIINGQKSTGNSKSPK-----GLKTPKTKMKQMTLLDMAKGTQKMT 451
 DB 1526 SLVSPAKTII----QTDNSLTSPKQDMPTPTTQINGSSPTTIS-----GVTGGDMNVL 1574

QY 452 RAPRNSGGTPRTSSKPHKHLPPAALHL-----IAYYKENKORE-----DK--RS 493
 DB

DB 1575 LQAKKSGARNSINDEHHHDVSIIDDDLSOLDSEITNVVEEEDNRLSADLEQLKLDKIVRA 1634

QY 494 ALSG--VISKTARLSS-----BDRA-----RLPEE-----LRSLVKRYELLEHKKRWASM 538
 DB 1635 SLNCKEALKSTNQLRAACFGQDRFWRRYWKLPKAGGIFIEALESQNDICDYHEALEAM 1694

QY 539 SBEQRKEYLKKREELKKLKEKAKERRE--KEMLERLE-----KQRYEDQ 583
 DB 1695 DDKDADENKENSENEKDVAAESSEQPMVEDESIITKLEDGPASDVGMPESNQNAHQDE 1754

QY 584 ELTKNLPAPRLV-----DTPGELPNTLFGDVAMVVFSLSCYSGLLLPDAQVP--- 631
 DB 1755 EDDDDDDVTEINKVEPEIVDLGDDDDAAPP-----LPKIEPRPE 1794

QY 632 ITAVSLMEALSADKGGFLYLNRLVLTLLQTLQDEIAEDYGELGMKLSBIPILTHSVSEL 691
 DB 1795 IKVSEMELMGPPP-----TTWISTKTDFEAEI-----KIPSPGILMPTT-- 1835

QY 692 VRLCLRRSDVBESSEGSDDDNKDSAA-----FEDNEVQDFLEKL----- 732
 DB 1836 --LNNNTNNNNNNNSDNCDKLETGLGLGQQQONFSQSVIKTDEVKEDDCCIIVSTSSV 1893

QY 733 -ETSEPP-----ELTSEE-----KLQILTA--LCHRIILTYVQDH----- 765
 DB 1894 DDTPKWFSIVRRVPLISELPAEEGSGVGGELQISYANQNCQAQL---QLQGHMPDLINN 1950

QY 766 -----METROOMGAELMKE--RLAVLKEENDKKRAEKKRKEAKKNKENGKVENGL 815
 DB 1951 MQYISIPMDCEKVDTSKLGNECIFSLSGLDKQMLAKVBEYKAHKVESK-----NGL 2002

QY 816 G-----KTRKKRIVKFEPQVDE-----AEDMISAVKSRLLA--IQAKKEREIQE-R 861
 DB 2003 GSPRRHHTKDEEQAKLKDKEIDTEMETDADDLAGKEKFFRLRSDVDPDTGGGVSEGT 2062

QY 862 EMVKVLROQAEEERIRKHAABAKAPQEGIAKAKLVMRRT-----PIGTDRNHNRYWL- 914
 DB 2063 DVPKPIELRDE-----ALSOAYYHNIAWMSLSVQTYPIPIFLPLSTPDPDHRLL 2114

QY 915 -----FSDEVPGLFIEKGWVHDSIDYRFNHCKDHTVSGDEDEDYCPRSKANLGNVSM 967
 DB 2115 EQVLAGPFFERVHGVTV----- 2131

QY 968 NTQHGATATEVAVETTPKQGNLWFLCDSQKELDELLNCLHPQGIRESQJLERLEKRYQD 1027
 DB 2132 -----PRQRYGWQLDDEQKLQQLLTLNPSGLRERLEQENLQR--- 2171

QY 1028 ITHSHLARKPNGLKSCDGNQELNLF--LRSDI-----IEVATRLQKGG 1070
 DB 2172 -----FLGLEQPLG---VNYKLKSIDFPPEEFMPDKKGMNPKVAKRYE--- 2213

QY 1071 LGYVSETSSEFARV-----ISILEKLD-----FGSCVIA 1099
 DB 2214 LALISLESLEDKVASASMLKNWOLPNRVESELTLDSQEDVTEEDFVSIIPMIRERID 2273

QY 1100 LQASVIKFLQCFMAPK-----QKRRKLQSDSDAK 1129
 DB 2274 LEANIERRYLYKPLPSQTDGAHLAVIAQNHQTTTQTONSASAAAYLLQWQQQQQQQLAQ 2333

QY 1130 TSEVDE-----EKKVVEEAKVASA----- 1148
 DB 2334 QQQQQQSGAGNSLNPSSFNERTMALAAAAAASGPNATGVANSVAVAGATPCESGSGE 2393

QY 1149 -----LEKKWKTAREAQTFSRMHVLLGLMDACIKWDM- 1181
 DB 2394 PMSGNASPASNCSDRDEKVEQIPKGLVQWRDAVRSRHTTAQLAMALYVLESVCAMDKSI 2453

QY 1182 ----- 1181
 DB 2454 MKAYATKNKSSKSSAKKQATPSKKQQQKNKKEQLTPNGKESKSAINKPENQAPL 2513

QY 1182 -----AENAR----- 1186
 DB 2514 SIKINLKALAQNQCCLLTKTPPILTKSPASPSHPNNTSDSDSDFGKRTKKKSGGKRRR 2573

```
QY 1187 -----CKVCKKGBDDKLILDCENKAFHLCLRALVEVDPGEWQCPA 1230
Db 2574 SANNTNSSKYSNLQNCQCTSGENEDKLLLCDCGCKGHTYCFKPKMDNIPDGWYCYE 2633
QY 1231 C-QPATARN-----SRGRNY-----1245
Db 2634 CYNKATNERKVICVGGHRSFVGKMYCYDCLPRAYHADCYIPLLKVPKRGWYCHGICISR 2693
QY 1246 -----TEESASEDEDE-----SDEEEEBEEEBEEDYEYVAG-----1279
Db 2694 APPPKRSAGTSGSSSKSRDRDRESGSAKRRSDNSKTPAMEHWQOQOMPLAGDSHH 2753
QY 1280 -----LRLRPKRTIG-----KHSV 1294
Db 2754 HHHQPPSLNSHDSMSNLPAGSESVGACGRNGRRKSTRSGFQLNNASALCSPAHSV 2813
QY 1295 -----IPPA-----1298
Db 2814 VSATNYDDQHANNVDSGSRFPAHLIPPSNNGTAALLLEDVPGANVMPGVYVTPVAA 2873
QY 1299 -----ARSGR-----1303
Db 2874 GNFSAGLINQAPVQAMPANVAMSPRAVTPTRTRTPPTPAPTTPPPPTPLMQASP 2933
QY 1304 -----RPGKKPHS-----1335
Db 2934 TATALHVNAQSPPOQOHAQMTMPPSPAIGVGTATNQMSPPPIINIAIQAQEKLEK 2993
QY 1336 R-----SSRROSLEQKEETLHKIVKFSWPPREPVRDEADYDYDVITHPMDFOIVQNK 1392
Db 2994 KEKHATKMLKELAVCKTLGEMELHEDSWPFLPNTWQFTYKIIKTFMDLSTIKKK 3053
QY 1393 CSCGSYRSVQEFLLTMKVFTNAEVYN 1419
Db 3054 LQDLSYKTRDFCDVDRQIFDNCMFEN 3080

RESULT 39
Q6BV90 PRELIMINARY; PRT; 1230 AA.
AC Q6BV90;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Debaromyces hansenii chromosome C of strain CBS767 of Debaromyces hansenii.
GN ORFNames=DEHA0C05082g;
OS Debaromyces hansenii CBS767.
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Debaromyces.
OX NCBI_TaxID=284592;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CBS767;
RG Cenolevures;
RA Dujon B., Sherman D., Fischer G., Durrens P., Casaregola S., Lafontaine I., de Montigny J., Marck C., Neuveglise C., Talla E., Goffard N., Frangeul L., Aigle M., Anthouard V., Babour A., Barbe V., Barnay S., Blanchin S., Beckerich J.M., Beyne E., Bleykasten C., Boisarame A., Boyer J., Cattolico L., Confaniolero F., de Daruvar A., Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi A., Hantraye F., Hennequin C., Jauniaux N., Joyet P., Kachouri R., Kerret A., Koszul R., Lemaire M., Lesur I., Ma L., Muller H., Nicaud J.M., Nikolski M., Oztas S., Ozier-Kalogeropoulos O., Pellenz S., Potier S., Richard G.F., Strub M.L., Suleau A., Swennene D., Tekala F., Wesolowski-Louvel M., Westhof E., Wirth B., Zentou-Weyer M., Zivanovic I., Bolotin-Fukuhara M., Thierry A., Bouchier C., Caudron B., Scarpelli C., Gallardin C., Weissenbach J., Wincker P., Souciet J.L.;
RT "Genome evolution in yeasts.";
RL Nature 430:35-44 (2004).
[2]
```

```
RP SEQUENCE FROM N.A.
RC STRAIN=CBS767;
RA Genoscope;
RL Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; CR382135; CAG85925.1; -.
DR InterPro; IPR004022; DDT.
DR Pfam; PF02791; DDT; 1.
SQ SEQUENCE 1230 AA; 142836 MW; 8302C13FE2D9C460 CRC64;

Query Match 4.1%; Score 329; DB 2; Length 1230;
Best Local Similarity 20.0%; Pred. No. 5.8e-06;
Matches 285; Conservative 203; Mismatches 488; Indels 452; Gaps 63;

QY 23 FTIPIHQAFRTREYERLERYSERIWTCKSTGSSQLTKAEWEEBOEVAELLKEEPPA 82
Db 26 FHIAATKEFWLTYDYLMQMDYIYHRKFVCEITGNSCLTFFQAIESEKTHEVEVRNPE 85
QY 83 WYELVLEVMVHHNTAS-LEKLVDTAWLEIMTKYAVEGECDFEVGKEMKLVKVIKHIPLE 141
Db 86 ALKEHILRFLOFNRIITRLDQLVDKVLVFKNDYPPGETI-----ILKGSIL-----131
QY 142 KVDEATEKKSADGACDSSPSKENSQIAQHQKKEVTVKDEDEGRRESINDRARRSPKL 201
Db 132 -----NGQTESGTR-----QRTGIKEKIQYS-----152
QY 202 PTLKKKGERKWAAPPKFLPHKYDVKLQNEBK--IISNVPADSLIRTERPPNKEIVRYFIRH 259
Db 153 -----NPSGLTLTYLVVRANDQKQAIVIN---EKISDRNFTKWLIIKAFIKL 198
QY 260 NALRAGTGENAPVWVEDELVKYSLPSKFSDFLLDPYKYMTLNPSTKRTKNTGSPDRKPSK 319
Db 199 TMTSRHK-VGAPVWVKNKFAKRYLNQKYPD---DLKHQFSSTPSGDIWYEG--DELPSA 252
QY 320 KSKTDNSSLSPKLPKLVCHVHLKK-----SLSGSPLKVKNSKN---SKSPEEHLEEMKK 371
Db 253 ALEVNGYRKLAKKPAASSKNLKKKALNSILDEPPVMTQEDDIKKR1PPH1IPDVL 312
QY 372 MMSPNKLTHTNPHI PKKGPAPKPGKHSKPKL-----AKGRSKGILN 413
Db 313 EVDEDAHKSS---PQPAYTSKLAVSLNLFNLOYSKPIDEVLA VPENAKSNWNHLIQ 369
QY 414 GQK---STGNSKSP-----KGLKTKPT-----433
Db 370 KHKEQETDTKLDPEMEDEIKELGVRSVYSIQDALETWFLNVVHSLKLDFTFTDFDIY 429
QY 434 -----KMKQMTLID---MAKGTQKTRAPRNSG-----GTPTSSKPKHK 470
Db 430 AMGWTHDQFKRIGRCELLDEIWCALGAIVSNEVPKNSKLKLDSDVIYGLQITLPAEDSY 489
QY 471 LPPAALHLIIAYYKENKORE---DKRSALSCVSKTARLLSSEDRARLPEELRSLVKRYE 527
Db 490 INPPS-----KODNENEIGLEERGSDDI---EENILKSEDDADNDESHSL-----E 534
QY 528 LLEHKRWASMSGEORKEYLKKREBELKKLKERERREKEMLERLEKRYEDQELTG 587
Db 535 IINGSKONAKVYKSEGEVKLDDENDGDDDEEBEEBEGNDENEDDELD-----EDREHNA 591
QY 588 KNLPAPRLVDPTEGLPNTLFGD-----VAMVVEFLSCVSGLL-----LPDAQPI 632
Db 592 YIVMYRNIRPWHRLRKRFRDGNWQCILLGLVLSIVEYVPAKPAIDKVIYRVLAPIDMPP 651
QY 633 TAVSLM-----EALSADKGGFLYLNRLVILATLQLODEIAEDYBELGKMLSIPLTLHSV 688
Db 652 TSSTVMNOFYNALDID-----LKL-QVLSILVLMNGTVQRNVIDECLDSSSTV-----699
QY 689 SELVRLCRRSDVQSESGSDTDNKSAAFEDEVQDFLEKLETSFEFFELTSEKLQI 748
Db 700 -----LRRNRL-----DN-----IRDYKVAVDLAQKL-----721
QY 749 LTALCHRLIMTVSVODHMETRQQMSAEIWLKRLAVLKEENDKKRAEKOKKEMEAKNKEN 808
Db 722 -----HVEIHH-----EKLMAKEMENAKS-KGSKSK---SKNKLK 752
```

QY 809 GKVENGLGKTRKKRIIVKFPQVDTEADMSIAVKGRRLLAIOAKKERETQEREMKVYLE 868
Db 753 GDSQTAVENTDAVKRRM---PRLNLQTFEM---TEQESKLAEDKGFQKCEQRKEALL- 805
QY 869 RQAEERIRKHAAAKAFOE-GIAKAKLVMTPTPIGTDRNHNRYLWFSDE-VPGLFIEK 926
Db 806 ---SIEKYQTKQKLEKQLEICQVRKL-----LQKRLYRNRYWFFNNGUPLTSHSNT 856
QY 927 GWYHDSIDYRNFHHCKDHTVSGDEDYCPRSKKNALGNKASMTNQHGTATVAVETTPKQ 986
Db 857 NDVDDND-----DENTA-----ANNNDDEDDND----- 881
QY 987 GQNLWFLCDSQKELDE--LNLCHLHPOGIRESQKLEKRYQDI--IHSHLARKPNLGL 1042
Db 882 -----DDNAVLDLETYLMKGLWQGPSNEDLRHFESNLEDIKYHNLYL----- 925
QY 1043 KSCDGNQELNFLR-----SDLE-----VATRLQKGGIGYVEETSE-----FE- 1081
Db 926 ---NYQOELNFYKCNNDQGEDNEDFKENKPEFKQNFVKLPLGFRSMSEMFGLSFEP 982
QY 1082 ---ARVISLEKLDGECVIALQASVKKFLQGFMAPKQRRKLQSDSA----- 1128
Db 983 NQVTKLLPDSKEDI---VIDHGLGLTVSSLVNLYLPLQKRVLEENPDLFSSSYRFDY 1039
QY 1129 KTEVDEEKQWBEAKVASALEKWKTAIREAQTFSRMHVILGMLDACIKWMSAENARCK 1188
Db 1040 KPEQV-----TKLIGLNPW-----GEREALLRKELL--VVKEAIVQSMARRKALS 1084
QY 1189 VCPKGGEDDKLILCDECNKAFHLCRLPALYEVDPGEWQCPACOPATARNRSGRYTSE 1248
Db 1085 LDQOETKLEL-----EHN-----IHQV-----TEKLSIGKDI--D 1114
QY 1249 SASSEDSDESEEEEEEEEDVEVAGLRPRKTIRGHVVP 1296
Db 1115 SISNLNDEAESEVQEEDEENED-----LRVGNKRLNK--VVP 1155

RESULT 40
USOL YEAST

ID USOL YEAST STANDARD; PRT; 1790 AA.
AC P25356;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Intracellular protein transport protein USOL.
GN Name=USOL; Synonyms=INTL; OrderedLocusNames=YDL058W;
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes;
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=X2180-1A;
RX MEDLINE=91185402; PubMed=2010462; DOI=10.1083/jcb.113.2.245;
RA Nakajima H., Hirata A., Ogawa Y., Yonehara T., Yoda K., Yamasaki M.;
RT "A cytoskeleton-related gene, usol, is required for intracellular
RT protein transport in Saccharomyces cerevisiae";
RN J. Cell Biol. 113:245-260(1991).
RN [2]
RP SEQUENCE OF 782-1790 FROM N.A.
RA Hostetter M.K., Herman D.J., Bendel C.M., McClellan M., Tao N.,
RA Kendrick K.E.;
RL Submitted (FEB-1993) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 1-8 FROM N.A.
RA Bai Y., Symington L.S.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Required for protein transport from the ER to the Golgi
CC complex.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic. Associated with intracellular
CC membranes. Probably present on vesicles operational between the er
CC and the Golgi complex.
CC -!- DOMAIN: The rodlike tail sequence is highly repetitive, composed

of an heptapeptide repeat pattern characteristic of alpha-helical
coiled coils. May form filamentous structures in the cell.
CC -!- SIMILARITY: Belongs to the VDP/USOL/YBL047C family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL; X54378; CAA38253.1; -
DR EMBL; L03188; AAB00143.1; -
DR EMBL; U53668; AAB66659.1; -
DR HSSP; P17676; I104.
DR GernOnline; 140300; -
DR SGD; S000002216; USOL.
DR InterPro; IPR008938; ARM.
DR InterPro; IPR002017; Spectrin.
DR InterPro; IPR006955; Usol_p115_C.
DR InterPro; IPR006953; Usol_p115_head.
DR Pfam; PF04871; Usol_p115_C; 1.
DR Pfam; PF04869; Usol_p115_head; 1.
KW Coiled coil; Cytoskeleton; Golgi stack; Protein transport; Transport.
FT DOMAIN 1 724
FT Coiled coil (Potential).
FT DOMAIN 725 1790
FT Charged (hyper-hydrophilic).
FT DOMAIN 465 487
FT DISPENSABLE FOR THE PROTEIN FUNCTION.
FT DOMAIN 991 1790
FT Asp/Glu-rich (acidic).
FT DOMAIN 1172 1786
FT G -> E (in Ref. 2).
FT CONFLICT 847 847
FT E -> K (in Ref. 2).
FT CONFLICT 924 924
FT V -> I (in Ref. 2).
FT CONFLICT 1253 1253
FT I -> V (in Ref. 2).
FT CONFLICT 1319 1319
FT N -> S (in Ref. 2).
FT CONFLICT 1461 1461
FT G -> S (in Ref. 2).
FT CONFLICT 1581 1581
FT I -> V (in Ref. 2).
FT CONFLICT 1600 1600
FT R -> S (in Ref. 2).
FT CONFLICT 1661 1661
FT D -> DEEDDEE (in Ref. 2).
FT CONFLICT 1772 1772
SQ SEQUENCE 1790 AA; 206424 MW; 6CE2B216E9FD4818 CRC64;

Query Match 4.1%; Score 325; DB 1; Length 1790;
Best Local Similarity 20.2%; Pred. No. 1.3e-05;
Matches 274; Conservative 222; Mismatches 471; Indels 388; Gaps 58;
QY 17 PCSEEPFPTPHQEARFRTREEYEARELRYSEIWTCKSTGSSQLTHKEAWE-----EE 69
Db 719 PDEEPISKI-----SPEVEKLQROCTKLKGBITSLOT--ETESTHENLTKLIALTNEH 771
QY 70 QEVAE-----LLKEEFPWYKLVLEVMVHNTA-SLEKLV-----DT 105
Db 772 KELDEKYQILNSSHSLKENFS-----ILETELKNVRDSLDEMTQLRDVLETKENQT 825
QY 106 AWLEIMTKYAVGEEC--DFEVGKMKLVKVIKIHLEKV-----DEBATEKKSDG 154
Db 826 ALLEYKSTIHQEDSIKLEKLETLISQKKAEDGINKMGKDLFALSREMVAEEN--- 882
QY 155 ACDSFSSDKENSQIAQDHQKKTETVVEDEGRRESINDPARRSPKLPFSLKKGKRWAP 214
Db 883 -CKNLQEKDKSN---VNHQETKSLKEDIAAKIT----- 913
QY 215 PKFLPHKYDVKLQNEDKLISNVPADSLIRTPPPNKEIVRYFIRHNALRAGTGENAPWV 274
Db 914 -----EIKAINENLEEMKIQCNLSKEKEHISKELVEYKSRQS----- 952
QY 275 EDELVKYSLPSKFSDFLLDPYKYMTLNPSTPKRKTGSPDRKPSKSKTDSNLSPLNP 334
Db 953 HDNLVAK--LTEKLS--LANNYKDMQMAENESLIKAV-----ESKNESIIQLSNLQN 1001
QY 335 KLWCHVHLKKSLSGSPKLVKVNKSNK-----SPEEHLEEMWQMSPNKLTNTHIPKGGP 390
Db 1002 KI-----DSMS-----QEKENFQIERGSIETKNIQLKKTISOLE-OTKEEIIKSDS 1047

```
QY 391 AKKPGKSHDKPLKAGKSGILNGKSTGNSKSPKKGLKTPKTKMKQMTLLDMAKTQKM 450
Db 1048 SDEYESQISLKEKLET-----ATTANDENVNKISLTKTRELEAEALAYKNL--- 1097
QY 451 TRAPNSGCTP-RTSKPHKHPALHAHIAIYKENDKREDKRSALSVCVISTARLLSS- 508
Db 1098 ----KNELETSETSEKALKEVENEHL-----KEEKIQLEKEA-----TETKQANSL 1143
QY 509 -----EDRALPSELVSLVQRYELLEHLKGRWAS-----MSEQRKYLKKBREL 554
Db 1144 RANLESLEKEHEDLAAQLKKYEQIANKERQYNEEISQLNDEITSTQGENESIKKNDL 1203
QY 555 KKKLKEKAKEREKEMLEKLEKQRYEDELTKGNLPAPRLVDTPEGUPNTLFGDVAMVV 614
Db 1204 -----EGEVKAKWSTSEOSNLKKSEIDALNLQIKELKKKNETWEASLESIKS-V 1253
QY 615 EFLSCVSGLLLPDAOVPIPAVSLME-----ALSADKGGFLYLNRLVILLQTLQDIAEDY 671
Db 1254 ESETVKIKELQDECNPKKEVSELEDKUKASEDKNS-KYLE-----LOKSEKIK 1302
QY 672 GELGKWLSEIPIULHSVSELVLCRRSDVQESGSDTDDNKDAAAFEDNEVDPELEK 731
Db 1303 BELDAKTETKIQLEKITNL-----SKAKESELSRLKKTSS--EERKNAEQLK 1353
QY 732 LETS-EFFELTSEKLOILTALCHRLIMTYSVODHMETRQOQSAELWKERLAVLKEENDK 790
Db 1354 LKNEIQIKNAQEKERKLNESGSTITQEYS--EKINTLE-----DELIRQNEEL 1403
QY 791 KRAE-KQKREKEMAKNKGKENGIGKTDKRIKVPQVDTBAEDMISAVKSR----- 845
Db 1404 KAKEIDNTRSELEKVSLSNDEL-----LEEKNTIK-----SLQDEILSYKDKITRN 1450
QY 846 ---RLIAIQAKEREIQS-----REMKVLEROAE--EERIKHKAHAFAFOEGIAK 893
Db 1451 DEKLLSIEDNRKDLSEKQLRAAQESKAKVEEGLKKELESSKEAELE-----K 1502
QY 894 AKLVMRRTIGTDNRHNRWLSDEVPGLFIEKGWVHDSIDYRFNHCKDHTVSGDEDYC 953
Db 1503 SKEMMKLESTIESNETE-----LKSSME-----TIKRSDEKL 1535
QY 954 PRSKANLGNKSNMTHQGTATEVAVETTPKQGNLWPLCDQSKELDELLNCLHPQGI 1013
Db 1536 EQSKKS--AEDIKNLQ-----EKSDLSIRNSEKDIKELSKLRIEAKS 1580
QY 1014 ESQKERLEKRYODIHSIHLARKNPLGKSCDGNQELLNFRSLDIEVATRLQGLGY 1073
Db 1581 GSEL-ETVQELNNAQEKIRINAEENTVLKS-----KLEDIE----- 1616
QY 1074 VEETSEFEARVTSLEKLDGFCVIALQASVIKFLQGFMAKPKRRKQLQSDSAKTEV 1133
Db 1617 -RELKDKQAEIKSNQSEKE-----LTSRLKELEQLDSTQQAQKSEEBERRAEVRKF 1668
QY 1134 DEEKQWVEEAKVASALE-KWKTAIREAQTFSR-----MHVLGLMDACIK 1177
Db 1669 QVEKQLDEK--AMLETYNDLVNKEQAKWKDEDVTKTTSQOEIEKLAKELD---- 1722
QY 1178 WMSAENARKVC--PKGSEDKLILCDENKAFHLFLRUPALYEVDPGEWCQACOPAT 1235
Db 1723 -NLKAENSKLKEANEDRSIDDLMLIVTDLDE----- 1753
QY 1236 ARNSRGRNVTESASEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDE 1270
Db 1754 --KNAYIRSKLDLGVIESSDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDE 1786
```

RESULT 41

```
Q6CAZ9 PRELIMINARY; PRT; 969 AA.
ID Q6CAZ9
AC Q6CAZ9
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Yarrowia lipolytica chromosome C of strain CLIB99 of Yarrowia
```

```
DE lipolytica (Fragment).
GN ORFNames=YALI0C3056g;
OS Yarrowia lipolytica CLIB99.
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Dipodascaceae; Yarrowia.
NCBI_TaxID=284591;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CLIB99;
RG Genolevures;
RA Dujon B., Sherman D., Fischer G., Durrens P., Casaregola S.,
RA Lafontaine I., de Montigny J., Marck C., Neuvéglise C., Tallia E.,
RA Goffard N., Frangoul L., Rigle M., Anthouard V., Babour A., Barbe V.,
RA Barnay S., Bianchin S., Beckerich J.M., Beyne E., Bleykasten C.,
RA Boisrons L., Boyer J., Cathodic C., Confanioleri F., de Daruvar A.,
RA Despons A., Fabre E., Fathhead C., Ferry-Dumazet H., Groppi A.,
RA Hantraye F., Hennequin C., Jauniaux N., Joyet P., Kachouri R.,
RA Kerrest A., Koszul R., Lemaire M., Lesur I., Ma L., Muller H.,
RA Nicaud J.M., Nikolski M., Oztas S., Ozier-Kalogeropoulos O.,
RA Pellenz S., Potier S., Richard G.F., Straub M.L., Suleau A.,
RA Swenne D., Tekala F., Wesolowski-Jouvel M., Westhof E., Wirth B.,
RA Zentou-Meyer M., Zivanovic I., Bolotin-Fukuhara M., Thierry A.,
RA Bouchier C., Caudron B., Scarpelli C., Gaillardin C., Weissenbach J.,
RA Wincker P., Souciet J.L.;
RT "Genome evolution in yeasts.";
RL Nature 430:35-44(2004).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CLIB99;
RG Genoscope;
RA Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.
RL EMBL, CR382129; CAG82483.1; -.
DR InterPro; IPR004022; DDT.
DR Pfam; PF02791; DDT; 1.
DR PROSITE; PS50827; DDT; 1.
FT NON TER
SQ SEQUENCE 969 AA; 112232 MW; 6CA55F880E6029BE CRC64;
```

```
Query Match 4.1%; Score 324.5; DB 2; Length 969;
Best Local Similarity 19.5%; Pred. No. 6.9e-06;
Matches 242; Conservative 176; Mismatches 396; Indels 427; Gaps 46;

QY 4 LLGRKPPFLVNLPGEE--PEFTTPTQEAFTREEYEARELYSERIWTCKSTGSSOLT 61
Db 2 LFKRAIHFPDIYVEEDDEVWIDATGEYFNVDYLTDMDFYKQKFICEVTGHSCLT 61
QY 62 HKEAWEEBQEAELLKEEFPAPYKELVLEMHVHTAS--LEKLVDPTAMLEIMTKYAVGEC 120
Db 62 LFEALKSELHSGSEIMDAFPENLKEPVLRRIOFTISRDLGLVDLTLYGEFKKDFPPE-- 119
QY 121 DFEVGKEMLVKVIKIHPLKVEDEATEKSDGACDSFSSDKENSSQIAQDHQKKEVW 180
Db 120 -----IVIAII-----ETAT----- 129
QY 181 KEDEGRESINDRARRSRPKLPTSLKKGCKWAPKPLPHKYDVVKLONEDIISNVPADS 240
Db 130 ----GATERVPAIIREKAKFNISITLDSGDVPAATT-----KYRLEAVSDPNKYVYCDFTQ 180
QY 241 LIRTERPNKEIVRYFIRHNALRAGTGEN---APMWVEDELVKYKSLPSKFSDFLLDPYK 297
Db 181 LTRDRKNFTVTLTKTFIKYSAKK---ENWNGAPWLVRPEYAEKYRIDQTPVPHL----- 231
QY 298 YMTLNPSTKRKNKTGSPDRKPSKSKTDNSSLSSPLNPKLWCHVHLKSLSGSLPKVNSK 357
Db 232 -----TQYRGNFSPE-----ELRKQRTQDRLKAQAR 258
QY 358 NSKSPPEHLE-EMMKMSPNKLHTNFIHPKKGPPAKPGKHSKPLKAKGRSKGI----- 411
Db 259 DNEKNDIILQKQKLNENKEK-----NAADALIRTLGSLGSKERDI 300
QY 412 ---LNGQKSTGNSKSPKKGLTKPTKMKQMTLLDMAKTQKMTAP-----RNSG 458
Db 301 QVPVSHQNGGNG-GAGLGAKTLATLASV-----TERRSPPKQVIREDLHLVVG 352
```

Qy	459	GTPTSSKPKH	-----LPPAAL-----HLIAYYKE-----	-----NKRED	490				
Db	353	VNP	-----KPVPHFDKLYPSDLTEVSLVWTFVINYKEVLVLDSTFYDDFDALRYEGESD	:	408				
Qy	491	KRSALSC	-VISKARTARLLSSDRARLPBRLSLVKCR-YELLEHKRWASMSSEORKEYLK	548					
Db	409	LLNEIHCALLISQVVGTSNDSLLVEYDDISDDDEEYDDEEEEEEEVKEQESEEVK	468						
Qy	549	KKR	-----EELKKLKEKAKEREKEMLERLEKQRYEDQ-----ELTKNLPA	592					
Db	469	EEEPGRTRSSRRQIKKEETEEAESEEEEEEEEEEEQEQKNSKSVSPGAENKANYAV	528						
Qy	593	FRLYDTPEGILPNTLFGDVAMVVEFLSCYSGLLLPDAQYPTAVSLMEALSADKGGFYLYN	652						
Db	529	FKGYDMKTRLQRRNFKD	-----GGW--Q	549					
Qy	653	RVLLVILLQTLQDSIAEDY--GELGM-KLSEIPLTLHSVSELVRLCLRRSDVQSESGSD	709						
Db	550	VILIGLLYNLYTEYPWEDFNGCLDVAAAKAPVSLASAD	-----	590					
Qy	710	TDDNKOSAAFEDNEVQDFLEKLETSBEFFLTSEKLIQILTALCHRLIMTYSVQDHMETR	769						
Db	591	-----RYFDLDFQVKVIQILCLHPTDAIRAYIEEC	624						
Qy	770	QQMSAELWGERLAVLKE	-----ENDKRAEKQKREMEAK-----NKEN	808					
Db	625	MGRVTELRDDKVDKQREYKIVSDDIRVLENEVKRIE---LEMEKKIPEDKDGFEKAE	680						
Qy	809	GKVENGLGKTDKRRKIVKFPQVDTEAEDMISAVKSRRLALAIQAKKEREIOEREMKV--	866						
Db	681	GNNGNESEPPSKS-----KTNDLSTLRK-----IDRQETLKEI	720						
Qy	867	-LERQAEERIRKHAAAEKAFOGIAKAKLWMRTPIGTDNRHNYLWLF-----	915						
Db	721	IAVINAKDEIR	-----FAD-----VNRLLLGKDRFNRTWFEKGIEDAGG	764					
Qy	916	--SDEVPLFTIEKGWVHDSIDYRPNHCKDHTVSGDEDYCPRSKANLGRNMSMTQHGT	973						
Db	765	SEEEBGTGTWRLWVQGP---FEEIK-FCLTGD--IPOSKR-----	801						
Qy	974	ATEVAVETTPKQGNLWFLCDSQKELDELLNCLHPQIGRESQLEKRYQDIHSIH	1033						
Db	802	-----EVEGEGYLEDYTKWGLEHPKELGELIKWLNQWGREHKLKCELLSVLKIKKASM-	856						
Qy	1034	LARKPNLGLSKDCNQBLNFLRSLDLTVATRLQKGLGVYEETSEPEARVLSLEKLKDF	1093						
Db	857	VSRLDLGREN-ELRKEDRKEIKDDEKVD-----ENTDEHQNGASSGDGKN	905						
Qy	1094	GECVIALQASVIKKFLOGFMAPKQKRKLQSEDSAKTEYVD	1134						
Db	906	GE-----TKEDKENEKQKENDVD	925						
RESULT 42									
Q62I78									
ID	Q62I78	PRELIMINARY;							
AC	Q62I78	PRT; 617 AA.							
DT	05-JUL-2004	(TrEMBLrel. 27, Created)							
DT	03-JUL-2004	(TrEMBLrel. 27, Last sequence update)							
DT	05-JUL-2004	(TrEMBLrel. 27, Last annotation update)							
DE	Putative DDT domain-containing protein.								
GN	Name=OJ1008.F08.22;								
OS	Oryza sativa (japonica cultivar-group).								
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;								
OC	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;								
OC	Ehrhartoideae; Oryzaceae; Oryza.								
OX	NCBI_TaxID=39947;								
RN	[1]								
RN	SEQUENCE FROM N.A.								
RP	Sasaki T., Matsumoto T., Yamamoto K.;								
RL	Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.								
DR	EMBL; AP004017; BAD15468.1; -.								

```

QY 934 DYRFNHCKDHTVSGDEYCPSPKANKLGNASMTQHGTAETVETTPKQGNLWFL 993
DB 524 -----SAD-----SK-----WG 532

QY 994 CDSOKELDELNCLPQIGRESQLEKRYQDIHSHILARK 1037
DB 533 YSTKEELVMSLNVKGLRERALKRQDLKLYSKISNALEKSK 576

RESULT 43
Q7LGT1 PRELIMINARY; PRT; 1268 AA.
AC Q7LGT1;
DT 05-JUL-2004 (Tremblrel. 27, Created)
DT 05-JUL-2004 (Tremblrel. 27, Last sequence update)
DT 05-JUL-2004 (Tremblrel. 27, Last annotation update)
DE US01 protein (Fragment).
GN Name=US01;
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OC NCBI_TaxID=4932;
RN [1]
RN [2]
RP SEQUENCE FROM N.A.
RA Bloeker H., Brandt P.;
RL Submitted (JUL-1996) to the EMBL/GenBank/DDBJ databases.
RP SEQUENCE FROM N.A.
RA MIPS;
RL Submitted (JUL-1996) to the EMBL/GenBank/DDBJ databases.
DR EMBL; 274105; CAA98620.1; -.
DR SGD; S00002216; US01.
DR GO; GO:0005737; C:cytoplasm; IEA.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0008565; F:protein transporter activity; IEA.
DR GO; GO:0006886; P:intracellular protein transport; IEA.
DR InterPro; IPR003900; KID_repeat.
DR InterPro; IPR002017; Spectrin.
DR InterPro; IPR006953; US01_p115_head.
DR Pfam; PF02524; KID; 4.
DR Pfam; PF04869; US01_p115_head; 1.
FT NON_TER 1
SQ SEQUENCE 1268 AA; 147462 MW; 5821B964F61E0AB4 CRC64;

Query Match 4.1%; Score 324; DB 2; Length 1268;
Best Local Similarity 20.2%; Pred. No. 9.9e-06;
Matches 274; Conservative 222; Mismatches 471; Indels 388; Gaps 58;

QY 17 PGCEPFTPTPTQEAFTREYEARELRYSERIWTCKSTGSSQLTHKEAW-----EE 69
DB 197 PDEEPINKI-----SFEVEKLQROCTKLKGEITSLQT--ETESTHENLTKLJALTNEH 249

QY 70 QBAVE-----LLKEFPAPWYKLVLEMVHNTA-SLEKLV-----DT 105
DB 250 KELDKEYQLNLSHSSLSKNFS-----LLETLNKVRSLDMLTQDRDLVETKQENQT 303

QY 106 AWLEIMTKYAVGEC--DPEVGKMKLVKVIKIHPLEKY-----DEBATEKSGDG 154
DB 304 ALLEYKSTTHKQEDSIKTLEKLGLETILSQKKAEDGINKMGKDLFALSREMQAVERN--- 360

QY 155 ACDSFSSDKENSQIAQDHQKQETVVKDEGRRESINDPARRSPKLPSTSLKKGKRWAP 214
DB 361 -CKNLQKEDKSN--VNHQKETSLKEDIKAAKIT----- 391

QY 215 PKFLPHKYDVKLQNEDKIISNPADSLIRTPPPKKEIIVRYFIRNALRAGTGENAPWV 274
DB 392 -----EIKAINLENEMKIQCNLSKEKEHISKELVEYKSRFS----- 430

QY 275 EDELVKYSLPSKFSDFLLDPYKYMTLNPSTKNTGSPDRKPSKSKTDNSLSPLNP 334
DB 431 HDNLVAK--LTEKLKS-LANNYKDMQAEENSLIKAV-----BESKNESIQLSNLQN 479

```

```

QY 335 KLWCHVHLKSLGSPKLVKNKSNK-----SPEEHLBEMMKOMSPNKLHTNFHPIKGGPP 390
DB 480 KI-----DMS-----QEKENFOIERGIEKNIQKLTISLE-QTKEEIISKSDS 525

QY 391 AKPKGKSDPLKAKGRSGILNGQKSTGNSPKKGLTKPTKMKQMTLDMAGTQOM 450
DB 526 SKDEYESQISLLKEKLET-----ATTANDENVNKISLTKTRELEAEALAAAYKNL--- 575

QY 451 TRAPNSGGTP-RTSSKPKHLPPAALHLIAYYKENDREKRSALSVCISKTARLAS- 508
DB 576 ----KNELETKLETSEKALKEVNEEHL-----KEEKTQLEKBA-----TETKOOLNSL 621

QY 509 ----EDRARLPEELRSLVQRYELLEHKKRWAS-----MSEORKEYLKKRREL 554
DB 622 RANLESLEKEHEDLAAQLKKEEQIANKERYNEEISQINDEITSTQENESIKKKNDEL 681

QY 555 KKLKEKAKERREKEMLERLEKQRYEDQELTKGNLPAPFLVDTPEGPLNTLFGDVAMV 614
DB 682 -----EGEVKAMKSTSEQSNLKKSEIDALNLQIKELKKKNETNEASLESIKS-V 731

QY 615 EFLCYSGLLLPDAQYPTAVSLME---ALSADKGGFLYANRVLVILLQTLLODIAEDY 671
DB 732 ESETVKIKELQDECNFKPEKVESELEDKIKASEDKNS-KYLE-----LOKSEKIK 780

QY 672 GELGMKLSIEIPLTLHSVSELVRLCLRRSDVOESEGSDTDNKKSAAPEDNEVQDEFLEK 731
DB 781 EELDAKTTTELKIQLEKYNL-----SKAKESESELSRLKKTSS--EERKNAEQLEK 831

QY 732 LETS-EFFELTSEEKLIQILTALCHRLMTYSVQDHMETRQQMSAELKWERLAVLKEENDK 790
DB 832 LKNEIQIKNOQAFKERKLNBSGSTITQEYS-EKINTLE-----DELIRLQENEL 881

QY 791 KRAE-KQKKEWAKENGKVENGLTKDRKRVKPEQVDTAEEDMISAVKSR----- 845
DB 882 KAKEIDNTRSELEKVSLSNDEL-----LEEKONTIK-----SLQDEILSYKDKITRN 928

QY 846 --RLIAIOAKKERETQ-----REMKVKLEQAE--EERIRKHAAAKAFQEGIAK 893
DB 929 DEKLISIERDNKRDLESLEKQIRAAQESKAKVEGLKKLEBESSKEAELE-----K 980

QY 894 AKLVNRTPIGTDNRNHNWYLFSDVEPGLFTIEKGWVHDSIDYRFNHHCKDHTVSGDEDYC 953
DB 981 SKEMMKLESTIESNETE-----LKSSME-----TIKRSDEKL 1013

QY 954 PRSKANLGNASMTQHGTAETVETTPKQGNLWFLCDSQKELDELNCLHPQGR 1013
DB 1014 EQSKS--AEEDIKVLQ-----EKSLISRIINESEKDI BELSKLRIEAKS 1058

QY 1014 ESOLKERLEKRYQDIHSHILARKPNLGLKSCDGNQELLNFRSLDIEVATRLQKGLGY 1073
DB 1059 GSEL-ETVQELNNAQEKIRINAEENTVLS-----KLBDIE----- 1094

QY 1074 VEETSEPEARVISLEKLDGCECVIALQASVIKKFLOGFMAPKQKRLQSDSASKEBEV 1133
DB 1095 -RELKDKQAEIKSNQEEKE-----LLTSRLKEQLSDSTQQAQSEERRAEVRKF 1146

QY 1134 DEEKQWTEAKVASALE-KWKTAREAQTFSR-----MHVLLGMLDACIK 1177
DB 1147 QVKSQLEK--AMILETKYNDLVNKEQAKWRDETVKTTDTSQSQEIKELAKELD----- 1200

QY 1178 WDMAENARCKVC--PKGEDDKLTLCDENKAFHLFCLRLPALYEVDPGEMQCPACQPAT 1235
DB 1201 -NLKAEKSLKEANDRSEIDDLMLLVTDLDE----- 1231

QY 1236 ARNSRGNYTEESASESDEDESEDESEEEEE 1270
DB 1232 --KNAKYKSLKDLGVEISSDEEDDEDEDEDEE 1264

RESULT 44
Q07380
ID Q07380 PRELIMINARY; PRT; 1790 AA.
AC Q07380; P89892;

```


DT 01-NOV-1996 (Tremblrel. 01, Created)
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
 DE US01 protein.
 GN Name=US01;
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RN SEQUENCE FROM N.A.
 RA Bloeker H., Brandt P.;
 RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RN SEQUENCE FROM N.A.
 RA MIPS;
 RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
 DR EMBL; 274106; CAA98621.1; .
 DR PIR; S67593; S67593.
 DR HSP; P01106; INKP.
 DR SGD; S00002216; US01.
 DR GO; GO:0005625; C:soluble fraction; IDA.
 DR GO; GO:0006888; P:ER to Golgi transport; IMP.
 DR GO; GO:0006463; P:protein complex assembly; IDA.
 DR InterPro; IPR008938; ARM.
 DR InterPro; IPR002017; Spectrin.
 DR InterPro; IPR006955; US01_p115_C.
 DR Pfam; PF04871; US01_p115_C; 1.
 DR Pfam; PF04869; US01_p115_head; 1.
 DR SEQUENCE 1790 AA; 206450 MW; 90062544F55A52EE CRC64;
 Query Match 4.1%; Score 324; DB 2; Length 1790;
 Best Local Similarity 20.2%; Pred. No. 1.5e-05;
 Matches 274; Conservative 222; Mismatches 471; Indels 388; Gaps 58;
 QY 17 PCEEPFTPIPTQEAFTREEYEALRYSERIWTCKSGSQLPHKEAWE-----EE 69
 DB 719 PDEEPINKI-----SFEVEKLRQCTKLKGBITSLOT--ETESTHENLTKLIALTNEH 771
 QY 70 QEVAS-----LLKEEPAWYKLVLEGVHNTA-SLEKLV-----DT 105
 DB 772 KELDEKYOILNSSHSLKENFS-----ILETELKNVRDSDMTQLRDVLTKDKENQT 825
 QY 106 AWLEIMTKYAVEEC--DFEVGKMKLVKVIKHPLEKV-----DEATEKKSGD 154
 DB 826 ALLEYKSTIHKQEDSIKLEKLETLISQKKAEDGINKMGKDLFALSREMQAVEEN--- 882
 QY 155 ACDSPPSKENSSQIAQHQKKEVTVVDEBGRRESINDRARRSPKPLTSKKGKRWAP 214
 DB 883 -CKNLQKEKDSN---VNHQKETSLEKEDIAAKIT----- 913
 QY 215 PKFLPHKYDVKLQNEKLIISNPADSLRTERPPNKEIVRYFIRHNALRAGTGENAPVV 274
 DB 914 -----EIKAINENLEEMKIQCNLSKEKHEHISLVEYKSRFQS----- 952
 QY 275 EDELVKYTSPLSKFSDFLDPYKYNLPNTPSKRKNKTSKDTNSSLSSPLNP 334
 DB 953 HDNLVAK--LTEKLK--LANNYKDNQAEENSLIKAV-----ESKNESGSIQLSNLQN 1001
 QY 335 KLMCHVHLKSLSGSPKLYKNSKNSK-----SPEEHLSEMKMKMSPNKLTNPHPKGPP 390
 DB 1002 KI-----DSMS-----QKENFQIERGSIETKNIQKKTISDLE-QTKEEIISKDS 1047
 QY 391 AKPKGKSDKPLKAKGRSGILNGQKSTGNSKSPKGLKTPKTKMONTLDMAGTKQM 450
 DB 1048 SKDEYESQISLKEKLET-----ATTANDENVNKISLTKTREELBAELAAAYKNL--- 1097
 QY 451 TRAPNSGGTP--RTSSKPKHLPPAALHLIAYKENKDRKRSALSVCISKATRLSS- 508
 DB 1098 -----KNELETKETSEKALKEVENEHL-----KEEKIOLKEKA-----TETKQOLNSL 1143
 QY 509 ----EDRARLPPELRSVLQKRYELLEHKRWAS-----MSEQRKEYLKKREEL 554

DB 1144 RANLESLEKHEDLAAQLKKYEEQIANKEROYNEEISQLNDEBITSQENESIKKNDL 1203
 QY 555 KKLLKEKAKERREKEMLEKOKRYEDELTKGKLPAPRLVDTPEGLPNTLFGVAVVV 614
 DB 1204 -----EGEVKAKMSTSEBQSNLKSEIDALMLQIKELKKKNETWEASLLEIKS-V 1253
 QY 615 EFLSCYGLLLPDAQYPIYAVSLMB--ALSADKGGFLYLNRLVILLQTLQDIAEDY 671
 DB 1254 ESETVKIKELQDECNFKVEKVESELEDKLKASEDKNS-KYLE-----LQSEKIK 1302
 QY 672 GELGKMLSEIPLTHSVSELVRLCLRRSDVQBESEGSTDDNKDAAAFEDNEVQEFLEK 731
 DB 1303 EELDAKTTTELKIQLEKITNL-----SKAKESESELSRLKKTSS--EERKNAEQLBK 1353
 QY 732 LETS-EFFELTSEEKLIQLTALCHRLMTYSVQDHMETROQMSAELKWKERLAVLEENDK 790
 DB 1354 LKNEIQIKNQAFERKULNKGSSITIOEYS--EKINTLE-----DELIRLQNEHL 1403
 QY 791 KRAE--KQKREMEAKNKENGKVGNGLTKDRKRIKVFEPQVDTEAEDMISAKVSR---- 845
 DB 1404 KAKEIDNTRSELEKVSLSNDEL-----LEEKONTIK-----SLQDEILSYKDKITRN 1450
 QY 846 --RLIAIOAKKEREIQE-----REMKVKLERQAE--EERIRKHAAAEEKAFQEGIAK 893
 DB 1451 DEKLLSIEDNKRDLESLEKQLRAAQESKAKVEEGLKLEBESSKEKAELE-----K 1502
 QY 894 AKLVMMRTPIGTDRNHNRYWLPFSDVPGLFTEKKGWHDSDYRFNHNCKDHTVSGDEDYC 953
 DB 1503 SKEMMKLESTIESNETB-----LKSSME-----TIRKSDEKL 1535
 QY 954 PRSKANLGNKSNMNTQHTGATEVAVETTTTPKQGNLWFLCDSQKELDELNLCLHPQGITR 1013
 DB 1536 BQSKKS--AEEDIKNLQH-----EKSDLSIRNESEKIDIELKSLRIEAKS 1580
 QY 1014 ESQKLERLEKYQDIHSHIHLARKPNLGLKSCDGNQELLNLFRLSDLIIEVATRLQGGGLY 1073
 DB 1581 GSEL-ETVKQELNNAQEKIRINAEENTVLKS-----KLEBIE----- 1616
 QY 1074 VEETSEPEARVISLEKLDGFCVIALQASVIKKFLQGFMAPKQRRKLOQSDSAKTEEV 1133
 DB 1617 -RELKQDKQAEKTSQSEKE-----LTSRLKELEQLDSTQQAQSEERERRAEVRKF 1668
 QY 1134 DEEKMTVEAKVASALE-KWKTAIREAQTFSR-----MHVLGLMLDACIK 1177
 DB 1669 QVEKSQLEK--AMLETKYNDLVNKEQAKWRDDETVKTTTDSQRQIEKLAKELD---- 1722
 QY 1178 WDMAENARCKVC--PKGEDDKLILCDECKNAFHLFCLRPALYVEVPDGEWQCPACQPAT 1235
 DB 1723 -NLKAENSKLKEANEDRSEIDDLMLLVTDLDE----- 1753
 QY 1236 ARRSRGNYTEESASESDESDSEDEDEDEDEEE 1270
 DB 1754 --KNAKYKSKULGLGVEISSDEDEDEDEDEEE 1786
 RESULT 45
 Q6FWEO
 ID Q6FWEO PRELIMINARY; PRT; 1980 AA.
 AC Q6FWEO;
 DT 05-JUL-2004 (Tremblrel. 27, Created)
 DT 05-JUL-2004 (Tremblrel. 27, Last sequence update)
 DT 05-JUL-2004 (Tremblrel. 27, Last annotation update)
 DE Candida glabrata strain CBS138 chromosome D complete sequence.
 GN ORFNames=CAGL0D009249;
 OS Candida glabrata CBS138.
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
 OX NCBI_TaxID=284593;
 RN [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=CBS138;
 RG Genolevures;

AC Q6BU09;
 DT 25-OCT-2004 (TrEMBLrel. 28, Created)
 DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
 DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
 DE Similar to sp|P25386 Saccharomyces cerevisiae YDL058w US01.
 GN ORFNames=DEHAOC09658g;
 OS Debaryomyces hansenii CBS767.
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Debaryomycetes.
 OX NCBI_TaxID=284592;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CB5767;
 RG Genolevures;
 RA Dujon B., Sherman D., Fischer G., Durrens P., Casaregola S.,
 RA Lafontaine I., de Montigny J., Marck C., Neuvéglise C., Talia E.,
 RA Goffard N., Frangoul L., Aigle M., Anthouard V., Babour A., Barbe V.,
 RA Barnay S., Blanchin S., Beckerich J.M., Beyne E., Bleykasten C.,
 RA Boisarane A., Boyer J., Cattolico L., Confaniolieri F., de Daruvar A.,
 RA Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi A.,
 RA Hantraye F., Hennequin C., Jauniaux N., Joyet P., Kachouri R.,
 RA Kerrest A., Koszul R., Lemaire M., Lesur I., Ma L., Muller H.,
 RA Nicaud J.M., Nikolski M., Oztas S., Ozier-Kalogeropoulos O.,
 RA Pellenz S., Potier S., Richard G.F., Straub M.L., Suleau A.,
 RA Swennene D., Tekia F., Wesolowski-Louvel M., Westhof E., Wirth B.,
 RA Zenlou-Meyer M., Zivanovic I., Bolotin-Fukuhara M., Thierry A.,
 RA Bouchier C., Caudron B., Scarpeilli C., Gailliardin C., Weissenbach J.,
 RA Winkler P., Souciet J.L.;
 RT "Genome evolution in yeasts.";
 RL Nature 430:35-44(2004).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CB5767;
 RA Genoscope;
 RL Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.
 DR EMBL; CR382135; CAG86127.1; -;
 DR GO; GO:0005737; C:cytoplasm; IEA.
 DR GO; GO:0036020; C:membrane; IEA.
 DR GO; GO:0008565; F:protein transporter activity; IEA.
 DR GO; GO:0006886; P:intracellular protein transport; IEA.
 DR InterPro; IPR008938; ARM.
 DR InterPro; IPR003357; HEAT.
 DR InterPro; IPR003900; KID repeat.
 DR InterPro; IPR006955; Uso1_p115_C.
 DR Pfam; PF02985; HEAT; 1.
 DR Pfam; PF02524; KID; 3.
 DR Pfam; PF04871; Uso1_p115_C; 1.
 DR Pfam; PF04869; Uso1_p115_head; 1.
 SQ SEQUENCE 2042 AA; 233328 MW; 959292DED8EE1158 CRC64;

Query Match 4.0%; Score 315.5; DB 2; Length 2042;
 Best Local Similarity 18.9%; Pred. No. 3.9e-05;
 Matches 274; Conservative 251; Mismatches 502; Indels 419; Gaps 56;

QY 24 TIPHTQEAFTREEYEAELRYSERIW-TCKSTGSSQTHKEAWEEBOEVAELLKEEPPA 82
 DB 819 TTKSLQATEKSKKEFETSGKYFKSLQELALKKFGVSDDLVK-SLRQKLENSQARQKAE 877
 QY 83 WYEKLVLEMVH-----HNTASLEKLVDTFAWLEI--MTKYAVGECDDPEVGKEKMLK 131
 DB 878 GINKMSRELFLHSKQKESDSNISKHEKKVDSLKLEIANITYK-----TFEARIEKLOK 930
 QY 132 VK-IVKIHPLEKVEATEKKSADGACSPSSDKENSSQIAQHQKKTWVKEDEGRRESI 190
 DB 931 ANDLFK---EKV-EDLNKIKESVSYNESHSKKS-----REMKEK-LDDVEATNEHL 977
 QY 191 NDRARRSPKLPSTLKKGERKWA-----PPKFLPHKYDVKLQNEDKIINVPADSLIRTER 246
 DB 978 MDKLSAASAFAQ-----EMKYAKTSDDKEIKYIEIKSKN-DEFNSLQLEFTLLKEEK 1030
 QY 247 PP-NKEIVRYFTRHNALRAGTGENAPWVVEDELVKYSLPSKFSDFLLDPYKMYTLNPS 305

DB 1031 EKINEEPESFKVKSSA-EVGKLNSS---INDLKSADSLDSNHSTI-----SNEFE 1077
 QY 306 KRKVTGSPDRKPSKSKSTDNSSLSP---NPKUWCHVHLKKSLSGSPFKVKNK-SNSK 361
 DB 1078 ELKEKLSQESKRYSNLKI EYDKLLESKCNK-----LKKOLE-----DIDNSKENAEB 1127
 QY 362 PEEHLEEMKMMSPNKLHTNFPHIPKGPAPKPKGKSHDKPLKAKGRSKGILNGQKSTGNS 421
 DB 1128 KQKSVB--ELINFKTKHSKVREBLEKSLDAKTGEYNDIAIEKLNKDKISIAS----- 1177
 QY 422 KSPKKGLKTPKTKMQLTMDMAKGTQKMTRAPNSGTPRTSSKPKHLPAPALHLIAY 481
 DB 1178 -----LK--ETHSKVSELD-----SG-----HSKLSQLEAANSRCLDET 1210
 QY 482 YKENKDREDKRSALSCVISTARLLSSDRARLPEELRSLVQKRYELLEHKRWASMSSEE 541
 DB 1211 EKQIKEHLESSNLADQISALEKV-----KGELEASINNAEOE-----SNK 1251
 QY 542 QRKEYLVKKREELKKLKEKAKERREKEMLERLEKOKRYEDQELTGKNLPAPFLVDTPEG 601
 DB 1252 SRESFEKEKALNQLNTNLEAKOKAKRDLVQEEKAIAEKEL----- 1295
 QY 602 LPNTLFGDVAMVVFSLSCYGLLLPDAQYPITAVS-LMEALSADKGGFLYLRNVLVLQ 660
 DB 1296 -----AKLQKILDNSKL-----ETEVSELKSDITTKFKDEHTIINEKLSIKTK 1338
 QY 661 TLLQ-DEIADYGEELGMKUSEIPLTHSVSELVRLCLRRSDVQEESEGST--DDNKDS 716
 DB 1339 ELSEKKDQIENQES---KLMDLAKSLDNEKILVK-----DLKEKKESELETRIKELND 1388
 QY 717 AAFEDN-----EVODEFLE-KLETSEFFELTSEBKLIQTALCHRIILMTYVSQDHMETRQ 770
 DB 1389 IAYASNSKEMQTKNENLETKTKKLDLSTLNKFTNETKILKDLISDHEVS----- 1441
 QY 771 QMSAELWKEVLAVLKEENDKKRAEKQKRKEMAKNKNKENGKVENGLGKTRDKRIVKFBPQ 830
 DB 1442 -----ISSLKVDLDKKVQVEKERNMLSENSETVIKEYG---DKIKELEKALGI 1487
 QY 831 VDTAEDMISAVKSRRLLAQA-----KKEREIOEREMKVKLQRQ 870
 DB 1488 AKTAHESKLNAMSKEKKLADSLKSYINNEFEIHKNSDNESTTKLTKEIESUNVKLENERK 1547
 QY 871 AEEERIRKHAATAEKAFOEGIAKAKLVMMRPTPIGTDNRHNNRYWLFSDVPGLFTEKGWVH 930
 DB 1548 LSTSKLSEREABLKSEETLKATGKQL-----EDEIEKLKKEKDVVS 1589
 QY 931 DSIDYR-----FNHCKDHTVSGDEDYCPRSKKANI-GKNASMTQ 970
 DB 1590 DNFKLKNLEFSTLEKDLASKVVEIKSINNVTESYKESDD--IKSKTKQLENDLEAAQK 1646
 QY 971 HGTATEVAVETTP-----KQGNLWFLCDSQKE-IDELLNCLHP 1009
 DB 1647 FGDKTKBELDTLNQKIBELKSVNSNTBETWTKLKESESSYAALDEQKKSISOELSALKS 1706
 QY 1010 QGIRESQLKERLEKYQDIHISI-----HLARKPNLGLKSCDG 1047
 DB 1707 SDKAASEMTKOLENELQTLKDDIEEKSRSKKELEKSTLTSTINELNKLDMKKELDS 1766
 QY 1048 NQELNLFRLSLIEVATRLQKGLGYVEETSEFARVISLEKLD-----FGECVIALQA 1102
 DB 1767 EKSVIEKLSAELKB-HSKLSADLKEYKEKFPQLEKEHEQLKKFPAEGNIGHKMKELKS 1825
 QY 1103 SV--IKKFLQGFMAPKQGRKQOSE-DSAKTEEVDEEKWVEAKVASALEKWKTAIREA 1159
 DB 1826 KLDLSQDLDTAAMDLSKIESLNQELLSSTTKTKDDEIKLTKOLESTOALKNNEKELKD 1885
 QY 1160 QTFPSRMVLLGMLDACIKWMSA--EN-ARCK----- 1188
 DB 1886 LNNSKENI-----STLKEDLNALKENETRULKQDLSEIKANEDSLNKELESLSKSQKNKE 1939
 QY 1189 -----VCPKKGBDDKLIL---CDECNKA 1208
 DB 1940 KDNIDFDSQKANFLESEKKLKEEVALKKLKEQSSTWVPSRSELDLMLMSDIDEHNKK 1999

```
QY 1209 FHLFCLRPALYVDPGEWQCPACQAPATARNRGRNYTEESASEDEDESDEEEEEE 1268
Db 2000 Y-----KKQLKALGQDISSEDSDEDDDDDDDDDDDE 2034
QY 1269 EEEEE 1274
Db 2035 DDDDD 2040

RESULT 48
Q25920
ID Q25920 PRELIMINARY; PRT; 1510 AA.
AC Q25920;
DT 01-NOV-1996 (TREMELrel. 01, Created)
DT 01-NOV-1996 (TREMELrel. 01, Last sequence update)
DT 01-WAR-2004 (TREMELrel. 26, Last annotation update)
DE Mature-parasite-infected erythrocyte surface antigen.
GN Name=MESA;
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=D10;
RX MEDLINE=92158014; PubMed=1741020; DOI=10.1016/0166-6851(92)90231-8;
RA Coppel R.;
RT "Repeat structures in a Plasmodium falciparum protein (MESA) that
   binds human erythrocyte protein 4.1.";
RL Mol. Biochem. Parasitol. 50:335-347(1992).
DR EMBL; M69183; AA29651.1; -.
DR HSP; Q57997; IMJH.
DR InterPro; IPR001623; DnaJ N.
DR Pfam; PF00226; DnaJ; 1.
DR Pfam; PF06513; DUF1103; 14.
DR SMART; SM00271; DnaJ; 1.
DR PROSITE; PS50076; DnaJ_2; 1.
SQ SEQUENCE 1510 AA; 177185 MW; 95C0F57FFF76EDA9 CRC64;

Query Match 3.9%; Score 309.5; DB 2; Length 1510;
Best Local Similarity 18.4%; Pred. No. 56-05; Indels 453; Gaps 66;
Matches 289; Conservative 260; Mismatches 570;

QY 28 TOEAFRTREYBARLERYSERIWTCKTSSQLT--HKEAWP-----BEQVAEALLKEEF 80
Db 246 SKETGESKETGESKETGESKETGESKETGESKETGESKETGESKETGESKETGESKETGE--SKET 303
QY 81 PAWYKLVLEWVHHNTASLEKLVDTAWLEIMTKYAVGECDFEVGKMKLVKVIKIHPL 140
Db 304 GESKETRIYE-----ETKYNKITSEPRETEENV----- 330
QY 141 EKVDEATEKKSDGACDPSKDKNSQIADHQKKEIVVKED-EGRR--ESINDRARRS 197
Db 331 -KITESKOREGN-KVSGPYENSENSVTSEETKLAENEKEGKLGENVNDGASEN 388
QY 198 ---PKLPTSLKKGGRKWAFFLPHKDYVKLQNEDKIISNPADSLIRTERPP----- 248
Db 369 SEDPKLLEQENGTKESSEET-----KDDKPENEKK-----ADNKKSKKKKSKFFQM 438
QY 249 -----NKEIVRYFIRNALRAGTGENAPWVVEDELVKKYSL-----PSKFSDFL 292
Db 439 LGCNPLCNKI-----ETDDEBETLVVKDDAKKKHFLREANTKNDNEKDKL 487
QY 293 L---DPYKYMTLNPSTKTKNTGSPDRKPSKSKTDNSLSPLNPKLCHVHLKSLSGS 349
Db 488 LGEQKEDVKEKNDQKQKVLGEGDKEDVKEKNDQKQKVLGEGDK-----EDVKEKNDGK 543
QY 350 PLKVKNS--KNSKSPHEHLEEMKMSPNKLHTNFHPKKGPAPKPKGHSDKPLKAGRS 408
Db 544 KDKVLGSEKTKQKEIKYKVKVKKCKKV-----KKGIKENTEGNDKVKG 590
QY 409 KGILNGQKSTGNSKSPKKGLTKPTKTKMKQMTLLDMAKGTQKWRAPRNSGGTPTRTSSKP 468
```

```
Db 591 PEIIIEEVKEIKQVEDGIKENDTEGN-----DKVKGPEIITE-----EVKEEIK 636
QY 469 KHLPPAALHLJAYYKEN-KDREDKRSALSCVISTKARTLSSSDRARLPBELRSLVKRYE 527
Db 637 KQVEDG-----IKENDTEGNDK-----VKGPEIITEEVKEIKQVEGEGIKENDT 681
QY 528 LLEHKKRWASMSBEQRKEYLKKREELKKKLEKAKERREKEMLERL-----EKQKRYEDQS 584
Db 682 EGNQKVKGPEIITEEVKEEIKQVEE--GVKENDTESKDKVIGQEIITEEVKEEIKQVE 738
QY 585 LTGKNLPAFLVDTPEGLPNTLFGDVAMVVEPLSCYSGLLLPDAQYPTITAVSLMEALSAD 644
Db 739 EKGKNENILEIKDIVIG-QEVIIEEVKKVIK-----KKVE 772
QY 645 KGGFLYLNRVLVILLQTLQDEIADYGBELGMLKSEIPLTLHSVSELVRLCLRRSDVOBE 704
Db 773 KG-----IKENHTESKDKVIGQEIIVEEVKEE-----EKQVBSG 807
QY 705 SEGSDTDDNKDSAAFED--NEVQDFLEKLETSFFELTSSEKL-QILTALCHRLMTY 760
Db 808 IKENDT-ESKDKVIGEEVIGKDVNEEGPENKD-----KVTQKEVKVEKVKVKVKRV 861
QY 761 SVQDHMETROQ--MSAELWKEKRLAVLKEENDKRAEKOKRKEAKENKENGKVENGLGKT 818
Db 862 KCRNNKNERKDNVIGKEIMKEDVNEKDTANKDKETIEQEKKE--EYKESEEVKEKEV-- 917
QY 819 DRKKRIVKFEPOVDTEAEDMSIAVKSRRLLAIQAKKEREIOEREMKVKLERQAEERIRK 878
Db 918 -KEKEEVKEKEVKEKEE-----VKEK-----EEVKEKDTSEKKEIEQEKKEVKEVK 966
QY 879 HKAATAKAFQEGIAKAKLVMRRTPIGTDNRNHRNYLWFSDEVPG--LFTIEKGWVHDSIDYR 936
Db 967 EKDTENK-----DKVIGQEIIEE----- 985
QY 937 FNHCKDHTVSGDEDYCPRSKANGKASWNTQHTATETAVETTTPKQGNLWFLCDS 996
Db 986 -----IKKEVKRVRKRN--NKNENKDNVIGQEIIMNEDVNEKDTANKDKVIEQ 1031
QY 997 OKELDELLNCLHPQGIRESQKLERLE-KRYODIHSIHLARKPNLGLSKCGN-----QE 1050
Db 1032 EKEKEEVK-----EKEVEKKEVEKVEKVEKVEKVEKVEKVEKVEKVEKVEKVEKVEK 1083
QY 1051 LLNLRSLDIEIATRLQKGLGYVEETSEFARVISLEKJDKDFGECVIALQASVKKFLQ 1110
Db 1084 IMN---EDVNEKDT-----ESKDKMIGKEVIEEVKE--EVKKRVNKEVKRV-- 1126
QY 1111 GFMAPKQKRKLQSEDSAKTEVEDE-----KQWVEAKVVASALEK-W 1152
Db 1127 ---NRNRNKRKRDVIEQEVSEVNEKDTKNNDKKIGKRVKPIDDKCKEREVQBITM 1181
QY 1153 KTAIREAOTFSR---MHVLLGLMDACIKWDMSE-----NARCKVCPKGGEDDKLIIC 1202
Db 1182 NEDVNEKDTSEKDKMIGKEVIEEVKEVEKRVNKRNRNRNKRKRDVIEQEVIS 1241
QY 1203 DECN-----KAFHLFCLRPALYVDPGEWQCPACQAPATARNRSGRNYTEESASEDS 1254
Db 1242 EEVNEKDTKNNDKKIGKRVKVP-----IDD-----CKKEREVQEESESESESESESES 1293
QY 1255 EDDSDSESESESESESESESESESESESESESESESESESESESESESESESESESESE 1314
Db 1294 EEESESESESESESESESESE----- 1320
QY 1315 SQPAPPVDDAEVDDELVLQTKRSRPSQSLQKCEIILH-----KIVKRVSWPPEPT 1369
Db 1321 SESESESESESESESESESESESESESESESESESESESESESESESESESESESE 1376
QY 1370 RDEADYDVDTHPMDFQTVQNKSCGYSRVQBFLLTDMQ-----VFTN-----AE 1416
Db 1377 EKQNEEYFD-----EEVILQSKHEHTLNTFPNALNEDFRRE 1413
QY 1417 VYNGCRGSHVLSVMKVTBOCLVLLHKLHPGHYPVRRKRRKPPDLRAE--DEGDSPEAVGQ 1475
Db 1417 VYNGCRGSHVLSVMKVTBOCLVLLHKLHPGHYPVRRKRRKPPDLRAE--DEGDSPEAVGQ 1475
```



```
Db 1404 GHLEEKOGSIRPEQKEDSADIQNTLET-----VNISDVNDFOI 1444
RESULT 50
Q9W6V0
ID Q9W6V0 PRELIMINARY; PRT; 3616 AA.
AC Q9W6V0;
DT 01-NOV-1999 (Tremblrel. 12, Created)
DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)
DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)
DE Hyperion protein, 419 kd isoform.
GN Name-hyperion;
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]_TaxID=9031;
RP SEQUENCE FROM N.A.
RC TISSUE=Retina;
RA Kemmer W.A., Schwarz U.;
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ131892; CAB40350.1; -.
DR HSP; Q60351; IGO3.
DR InterPro; IPR005479; Cphg_synth_L_D2.
DR PROSITE; PS00867; CPSASE_2; UNKOWN 1.
SQ SEQUENCE 3616 AA; 418999 MW; 6D3BB330E6BCA92A CRC64;

Query Match 3.8%; Score 306; DB 2; Length 3616;
Best Local Similarity 18.9%; Pred. No. 0.00019;
Matches 351; Conservative 294; Mismatches 680; Indels 530; Gaps 78;

QY 27 HTQEAFTREYEARELRYSERIWTCKSTGSSQLTHKEAWEEQVVAELLKEEFPWYK 86
Db 794 HLOQVNTLKSQADLOLQWELQALCLSLVYS--AHVD-----QVREHLKAEKSELCS 845
QY 87 LVLEMVHNTASLEKLVDTAWLEIMTKYAVGECDPEVGEKEMLVKVIKIHPLKVDDE 146
Db 846 LKEELIRSHLQEMNDLKKVHOMELQTMKI--QPADDEVSSQRL-----IEKLEA 894
QY 147 ATEKS---DGACDPSDKENSSQIAQHQKQKTVKDEGRSINDRARRSPKPLPT 203
Db 895 ITECSRITQIFCDWN---EKSTVIETGDEQDIFSRPAAEKENLMDLTHRHGQIA 951
QY 204 SLKGERKWAAPKFLPHKY-----DVKLQNKDKLISNVPADSLRTERP 247
Db 952 MOEHNEA-----LLHKILEYNRKALQTLQVCKQVKELWLSSEP-----ERR 997
QY 248 PNKEIVRYPIRHNLRAAGTG-----ENAPWVVEDELVKKYS--LPSKFSDF 291
Db 998 KEEESNCLETRKESLHASSHNLMWPKIHLCWKEMENVKTLQEEQHAQVEHLRSYFQQ 1057
QY 292 LLDPKYMTLNPSTRKNTGSPDRPKSKKTDNSLSPLNPKWCHVHLKKSLSGSP 351
Db 1058 LKDSVRYTAETIAHLQ-----DKLRSEVSDYTSIAESQMK-----LKESPR--- 1101
QY 352 KVENSKN-----SKSPEHLE----- 367
Db 1102 KYKTENPHQGEDVALELANIEMKNDLDVVOLLEKQYERLEEEIAKIVMSVAFK 1161
QY 368 --EMMKMSPNKLTNFIHPK-----GP--PAKPGKHS-----KP 401
Db 1162 QSELRIARQKKEETQTOIEHQGMGFEMRECSIEVDGGLPKKATEKSKHEELSKLCE 1221
QY 402 LKAKGRSGILNQKSTGNSKPKGLKTPKTKMQMTLLDMAKTQKMTAPRN--SGG 459
Db 1222 LSESEGMGLL--GEQLCSNSQSGYVLRQTRESVTSERLFSHVQGHATAETLCNNVAA 1280
QY 460 TPRTSSK---PHKLPAPALHIAYYKEN-----KDRDKRSALSCVIS 500
Db 1281 DMETSSQLLIYERLEDMRQELVRQYQBHQQAATELLROGHMQOMERQENOEQLIALES 1340

501 KTAFL-----LSSEDRARLPPELRSLVKRY-----ELLEHKRWASSEEQ 542
Db 1341 LKQLAQRVSMENDNLAERERMLLEELSKLRQHPVPGKERLFCFLQNSSTOTENEDQ 1400
QY 543 R--KEYLKKRELEK-----KLEKAKERREKEMLEKLEKQKRYEDQ----- 583
Db 1401 NDVREQISEDEDEGRKPDEVSSALLSKERHVVQKANEKMLKMLLELVVTKTVAMEETGRH 1460
QY 584 -----ELTGKNLPAFRLVDTPEGLPNTLFGDVMVV-----EFLSCYSGLLLPDAQYPI 632
Db 1461 VLVLDRSGKVQPS-----KPAWGTEAEDSVKPSIHVGYEKESCSYHSGSGDDDDITM 1515
QY 633 TAVSLMEAL-----SADKGGFL-YLNRVLVILLQTLQ-----DRIADYGBELGMLKSLPIPT 684
Db 1516 WSGTAEBGLLSQHLAESGVDELDPENELVNLSSRLQAAVEKLEALAINETSQLEHAKIT 1575
QY 685 LHSVSELVRLCLRRSD-----VQSESGSTDDNK--DAAAFEDNEVDQEFLEKLETS 735
Db 1576 ---QTELMRESFKQEEATEFIRYQEBLQERLSEETKAREQLALELNKAESLIDGYADEK 1632
QY 736 EPEFELTSEKLIQILTALCHRLIMTYSVQDHMETROQMSAE----- 775
Db 1633 AFLEKQLEKIDVIDHLEQELCTGNKQLEAEQOIQEKEKELLARQKDMRADGPVE 1692
QY 776 -----LWKERLAVL-----KEENDKXR-----AEKQKRKEMEAR 804
Db 1693 QQLLEETEKLMEKTEVQQAKEYDDLQKQVKVLEIDLEEQVSRFIELEQEKNAELMDL 1752
QY 805 NKENGKVENGLKT-----DR-----KRIVFEPOV-----DTBA 835
Db 1753 RQNOALEKLEKTRKFLDEQAVDREHERDFQEBIQLEQLKVPQSPQVNEHQSRV 1812
QY 836 EDMISAVKSR-----RLLAIOAKKEREIQER----- 861
Db 1813 EQLTNHLEKTKCELSKELSLQORDIQERNEETEKLECRILEEQALIISADNLQKVE 1872
QY 862 -----EMVKLERQAEERI--RKHAAAE-----KAFQEGIAK-----A 894
Db 1873 ERKQGTIIVKGELPLEIQLAERAVDRKEKVTNLEBQLEQFRELKNEEVOQLHM 1932
QY 895 KLVMR-----TPIGTRNHNRYWLFSD--EVPGLFIEKGWVHDSIDYRPNHHCCKHTVSG 948
Db 1933 QLEIQRKESTHLOLEQENK--LFDKEMELGLAIOK-----SEDGTIKDH--HLVAG 1982
QY 949 DEDYCPKCANLGNKSMNTQHGATVAVETTPK--OGONLWFLCDSQKELDELNC 1006
Db 1983 KLAHMQKEQSID--NLHQIAKLOQLEGGTUNKIFEEQN-----EHIRELEAQVBC 2034
QY 1007 LHQOGIRE-----SQLKERLEKRYQDIHSIHLARKPNLGLKSCDGNQELLNLRSD 1058
Db 2035 LKSDQERVKKNDDEIEQLNDVIDKLQELANIEQIPSDITAFHEDADSLKHTLETVLAE 2094
QY 1059 LIEVATRLQKGLGVVESTSEFARVILEKLDKPGECVIALQASVKKFLOGFNAPKQK 1118
Db 2095 KEALEKQVESINLEASQTKNELETKELENNQLKQBINILRKEREVEVWEVKYKGLM--KGD 2152
QY 1119 RRLKQSDASAKTEEVDEEK---KXVEBAKVASALEKWKTAIREAQTSRMHVHLGMLDA 1174
Db 2153 REKTEVGNRRKTEKVDGSCWTEMLDQTLQURSSDENTRV-----TISKMEVQLOQLQA 2206
QY 1175 CIKMDMSAENARC-----KVCPPKKGEDDKLILCDCKNAKPHLFCRLPALVEVPDGEWQC 1228
Db 2207 CIK---EKDSELCSYNEIKDLKEQKAEKRMWGNK-----ILEKALVE-----KV 2251
QY 1229 PACQATARRNS---RGRNYTE--ESASDSED--DESDEEBEEEEEBEEDVEVAGLRIR 1283
Db 2252 AAALVSQQLNAVQEGKFLQEIQAKSKCLSEASKHPQTEGNDNTTENEVESEVSLTQR 2311
QY 1284 PRKTIIRGHKSVIPPAARSGRPGKKPHSTRSQPKAPPVDDAEVDLVLQTKSRSSRQSL 1343
Db 2312 LREV-----EQQLANTHSLLEL-----EKENVKVTQQESKLKEERLL 2348
QY 1344 ELQK-CEBILHK-----IVKYRFSWFPFRPVRDEAEYD-----DVITHPMDFQ 1387
```


GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 9, 2005, 14:24:09 ; Search time 103.022 Seconds
(without alignments)
4877.138 Million cell updates/sec

Title: US-10-702-148-27

Perfect score: 7967

Sequence: 1 MAPLGRKPPFLVPLPGEE.....NCFMMLVNTQFCMALTDVT 1527

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1391452 seqs, 329044822 residues

Total number of hits satisfying chosen parameters: 1391452

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 65 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
- 17: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
- 18: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pep.*
- 19: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pep.*
- 20: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	7967	100.0	1527	9 US-09-839-479-27	Sequence 27, Appl
2	7967	100.0	1527	15 US-10-376-537-27	Sequence 27, Appl
3	7967	100.0	1527	15 US-10-702-148-27	Sequence 27, Appl
4	7955	99.8	1531	9 US-09-839-479-29	Sequence 29, Appl
5	7955	99.8	1531	15 US-10-376-537-29	Sequence 29, Appl
6	7955	99.8	1531	15 US-10-702-148-29	Sequence 29, Appl
7	7888	99.0	1525	9 US-09-839-479-68	Sequence 68, Appl
8	7888	99.0	1525	15 US-10-376-537-68	Sequence 68, Appl
9	7888	99.0	1525	15 US-10-702-148-68	Sequence 68, Appl
10	2889	36.3	572	9 US-09-864-761-34546	Sequence 34546, A
11	2887	35.7	560	9 US-09-864-761-37938	Sequence 37938, A
12	929.5	11.7	1674	9 US-09-839-479-1	Sequence 1, Appl
13	929.5	11.7	1674	15 US-10-376-537-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1

- US-09-839-479-27
- Sequence 27, Application US/09839479
- Publication No. US20020039779A1
- GENERAL INFORMATION:
- APPLICANT: Jones, Michael H.
- TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
- FILE REFERENCE: 06501-042002
- CURRENT APPLICATION NUMBER: US/09/839,479
- CURRENT FILING DATE: 2001-04-20
- PRIOR APPLICATION NUMBER: US 09/418,710
- PRIOR FILING DATE: 1999-10-15
- PRIOR APPLICATION NUMBER: PCT/JP98/01783
- PRIOR FILING DATE: 1998-04-17
- PRIOR APPLICATION NUMBER: JP 9/310027
- PRIOR FILING DATE: 1997-10-24

14	929.5	11.7	1674	15	US-10-702-148-1	Sequence 1, Appl
15	918	11.5	1673	9	US-09-839-479-69	Sequence 69, Appl
16	918	11.5	1673	15	US-10-376-537-70	Sequence 70, Appl
17	918	11.5	1673	15	US-10-702-148-69	Sequence 69, Appl
18	712	8.9	141	9	US-09-764-864-1134	Sequence 1134, Ap
19	605	7.6	1586	13	US-10-087-192-663	Sequence 663, App
20	540.5	6.8	1873	13	US-10-087-192-666	Sequence 666, App
21	532	6.7	1972	9	US-09-839-479-21	Sequence 21, Appl
22	532	6.7	1972	15	US-10-376-537-21	Sequence 21, Appl
23	532	6.7	1972	15	US-10-702-148-21	Sequence 21, Appl
24	527.5	6.6	1876	9	US-09-839-479-70	Sequence 70, Appl
25	527.5	6.6	1876	15	US-10-376-537-71	Sequence 71, Appl
26	527.5	6.6	1876	15	US-10-702-148-70	Sequence 70, Appl
27	521.5	6.5	1878	9	US-09-839-479-13	Sequence 13, Appl
28	521.5	6.5	1878	15	US-10-376-537-13	Sequence 13, Appl
29	521.5	6.5	1878	15	US-10-702-148-13	Sequence 13, Appl
30	508.5	6.4	1969	9	US-09-839-479-71	Sequence 71, Appl
31	508.5	6.4	1969	15	US-10-376-537-72	Sequence 72, Appl
32	508.5	6.4	1969	15	US-10-702-148-71	Sequence 71, Appl
33	341	4.3	65	9	US-09-864-761-34543	Sequence 34543, A
34	338	4.2	69	9	US-09-864-761-34544	Sequence 34544, A
35	324	4.1	1790	15	US-10-369-493-1586	Sequence 1586, Ap
36	305.5	3.8	1327	16	US-10-408-765A-1215	Sequence 1215, Ap
37	304	3.8	1400	9	US-09-764-176-7	Sequence 254, App
38	299	3.8	1400	9	US-09-764-176-7	Sequence 7, Appl
39	296	3.7	1398	15	US-10-094-466-30	Sequence 30, Appl
40	296	3.7	2375	16	US-10-408-765A-277	Sequence 277, App
41	295	3.7	1827	15	US-10-369-493-5368	Sequence 5368, Ap
42	295	3.7	1881	14	US-10-032-585-7646	Sequence 7646, Ap
43	293	3.7	2492	16	US-10-697-526-2	Sequence 2, Appl
44	292.5	3.7	1972	14	US-10-171-311-162	Sequence 162, App
45	292.5	3.7	1972	15	US-10-341-434-103	Sequence 103, App
46	292.5	3.7	1979	10	US-09-927-597-4	Sequence 4, Appl
47	284	3.6	58	14	US-10-029-386-30379	Sequence 30379, A
48	281	3.5	65	9	US-09-839-479-67	Sequence 67, Appl
49	281	3.5	65	15	US-10-376-537-68	Sequence 68, Appl
50	281	3.5	65	15	US-10-702-148-67	Sequence 67, Appl
51	279	3.5	1819	15	US-10-335-977-7981	Sequence 7981, Ap
52	279	3.5	1820	15	US-10-335-977-7982	Sequence 7982, Ap
53	278	3.5	1938	14	US-10-171-311-164	Sequence 164, App
54	278	3.5	1945	10	US-09-927-597-2	Sequence 2, Appl
55	277.5	3.5	2020	15	US-10-369-493-5128	Sequence 5128, Ap
56	277.5	3.5	2020	15	US-10-369-493-5129	Sequence 5129, Ap
57	277	3.5	59	9	US-09-839-479-66	Sequence 66, Appl
58	277	3.5	59	15	US-10-376-537-67	Sequence 67, Appl
59	277	3.5	59	15	US-10-702-148-66	Sequence 66, Appl
60	276	3.5	1927	15	US-10-282-122A-58729	Sequence 58729, A
61	275.5	3.5	2503	16	US-10-828-985A-11	Sequence 11, Appl
62	275.5	3.5	2543	16	US-10-828-985A-9	Sequence 9, Appl
63	275.5	3.5	2568	16	US-10-828-985A-7	Sequence 7, Appl
64	272	3.4	149	14	US-10-195-730-317	Sequence 317, App
65	272	3.4	149	16	US-10-799-747-317	Sequence 317, App

```
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 27
; LENGTH: 1527
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-839-479-27

Query Match      100.0%; Score 7967; DB 9; Length 1527;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1527; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAPLLGRKPPFVLPVLPDGEPPFTTPTQEQAPRTREYEARELRYSEIRIWTCKSGSSQL 60
DB 1 MAPLLGRKPPFVLPVLPDGEPPFTTPTQEQAPRTREYEARELRYSEIRIWTCKSGSSQL 60

QY 61 THKEAWEEQEVAEILLKEFPFAWYKLVLEWVHHNTASILEKLVDTAWLEIMTKYAVGEEC 120
DB 61 THKEAWEEQEVAEILLKEFPFAWYKLVLEWVHHNTASILEKLVDTAWLEIMTKYAVGEEC 120

QY 121 DFEVGCKEMLVKVIKIHPLKXVDEATEKKSDDGACDSSPSDKENSSQIAQDHQKKEIVV 180
DB 121 DFEVGCKEMLVKVIKIHPLKXVDEATEKKSDDGACDSSPSDKENSSQIAQDHQKKEIVV 180

QY 181 KEDGRRSINDRARRSPKPLPTSLKGERKWAPPKFLPHKYDVVKLQNEDKIISNVPADS 240
DB 181 KEDGRRSINDRARRSPKPLPTSLKGERKWAPPKFLPHKYDVVKLQNEDKIISNVPADS 240

QY 241 LIRTERPNNKEIVRFIRHNALRAGTGENAPWVDELVKYSPLSKFSDFLLDPKYMT 300
DB 241 LIRTERPNNKEIVRFIRHNALRAGTGENAPWVDELVKYSPLSKFSDFLLDPKYMT 300

QY 301 LNPSTKRNTGSPDRKPKSKSTDNSLSPLNPKLWCHVHLKSLGSLGSLKVKVNSKNSK 360
DB 301 LNPSTKRNTGSPDRKPKSKSTDNSLSPLNPKLWCHVHLKSLGSLGSLKVKVNSKNSK 360

QY 361 SPEEHLEEMKMSNKLHTNPHIPKGGPPAKPKGKSDKPKLAKGRSGKILNGOKSTGN 420
DB 361 SPEEHLEEMKMSNKLHTNPHIPKGGPPAKPKGKSDKPKLAKGRSGKILNGOKSTGN 420

QY 421 SKSPKKGKLTPTKTKWQMTLLDMAGTKMTAPNSGTPRTSSKPKHKLPPAALHLIA 480
DB 421 SKSPKKGKLTPTKTKWQMTLLDMAGTKMTAPNSGTPRTSSKPKHKLPPAALHLIA 480

QY 481 YYKENKDREDKESALSCVISKTARLLSSDRARLPEELRSVLQKRYEYELLEHKRWASMS 540
DB 481 YYKENKDREDKESALSCVISKTARLLSSDRARLPEELRSVLQKRYEYELLEHKRWASMS 540

QY 541 EQRKYLKKGREELKKLKEKAKERREKEMLERLEKQKRYEDQELTGKNLPAFRLVDTP 600
DB 541 EQRKYLKKGREELKKLKEKAKERREKEMLERLEKQKRYEDQELTGKNLPAFRLVDTP 600

QY 601 GLPNTLFGDVAMVVFSLCYSGLLPDAQYPTTAVSLMEALSADKGGFLYLNRLVILLQ 660
DB 601 GLPNTLFGDVAMVVFSLCYSGLLPDAQYPTTAVSLMEALSADKGGFLYLNRLVILLQ 660

QY 661 TLLQDEIAEDYGELGKMLSEIPLTLHVSSELVRLCLRRSDVQEESEGSDTDNKKDSAAFE 720
DB 661 TLLQDEIAEDYGELGKMLSEIPLTLHVSSELVRLCLRRSDVQEESEGSDTDNKKDSAAFE 720

QY 721 DNEVQDEFLEKLETSEFPFELTSEKLIQTALCHRLMTYSVDHMETRQMSABELWKR 780
DB 721 DNEVQDEFLEKLETSEFPFELTSEKLIQTALCHRLMTYSVDHMETRQMSABELWKR 780

QY 781 LAVLKEENDKKAERQKQKEMAKKENGKVGNGKTDKRRKIVKFEPOVDTEADMIS 840
DB 781 LAVLKEENDKKAERQKQKEMAKKENGKVGNGKTDKRRKIVKFEPOVDTEADMIS 840

QY 841 AVKSRLLAIQAKKEREIQEREMKVKLERQAEERIRKHAAAEKAFQEGIAKAKLVMR 900
DB 841 AVKSRLLAIQAKKEREIQEREMKVKLERQAEERIRKHAAAEKAFQEGIAKAKLVMR 900

QY 901 TPICTDRNHNRYMLFSDVPGLFIEKGWVHDSIDYRFNHHCKDHTVSGDEYCPSSKAN 960
DB 901 TPICTDRNHNRYMLFSDVPGLFIEKGWVHDSIDYRFNHHCKDHTVSGDEYCPSSKAN 960

QY 961 LGKNASMTQHGTAATEVAVETTPKQGNLWFLCDQSOKELDELNCLHPQGIRESQK 1020
DB 961 LGKNASMTQHGTAATEVAVETTPKQGNLWFLCDQSOKELDELNCLHPQGIRESQK 1020

QY 1021 LEKRYQDIHSHILARKPNLGLKSCDGNQELLNLFURSOLIIVATRLQKGLGYVEETSEF 1080
DB 1021 LEKRYQDIHSHILARKPNLGLKSCDGNQELLNLFURSOLIIVATRLQKGLGYVEETSEF 1080

QY 1081 EARVTSLEKLDKDFGECVIALQASVKKFLOGFMAPKOKRKLQSEDSAKTEVEDEKMW 1140
DB 1081 EARVTSLEKLDKDFGECVIALQASVKKFLOGFMAPKOKRKLQSEDSAKTEVEDEKMW 1140

QY 1141 BEAKVASALEKWKTAIREAQTFSRMHVLLGMLDACIKWMSAENARCKVCPKKGDDKLI 1200
DB 1141 BEAKVASALEKWKTAIREAQTFSRMHVLLGMLDACIKWMSAENARCKVCPKKGDDKLI 1200

QY 1201 LCDECNKAFHLCFLRPALYEVDPGEWQCPACOPATARNRSGRNYTESSESDESDSD 1260
DB 1201 LCDECNKAFHLCFLRPALYEVDPGEWQCPACOPATARNRSGRNYTESSESDESDSD 1260

QY 1261 EEEEEEEEEEDYEVAGLRPRKTIKRGKSVIPPAARSGRRPKKPHSTRSOPKAP 1320
DB 1261 EEEEEEEEEEDYEVAGLRPRKTIKRGKSVIPPAARSGRRPKKPHSTRSOPKAP 1320

QY 1321 PVDDAEVDDELVLQTKRSRRSLELQKCEEILHKIVKRYFSWPPREPVTREAEYDVI 1380
DB 1321 PVDDAEVDDELVLQTKRSRRSLELQKCEEILHKIVKRYFSWPPREPVTREAEYDVI 1380

QY 1381 THPMDPQTVQNKSCGYSYRSVOEFLTDMKVFTNAEVNCRGSHVLSWVKTEQCLVLL 1440
DB 1381 THPMDPQTVQNKSCGYSYRSVOEFLTDMKVFTNAEVNCRGSHVLSWVKTEQCLVLL 1440

QY 1441 HKHLPGHYPYVRKPKFPDLAEDEGSEPAVCGSRDEDRSRREAEIOEWLQDTSLSYA 1500
DB 1441 HKHLPGHYPYVRKPKFPDLAEDEGSEPAVCGSRDEDRSRREAEIOEWLQDTSLSYA 1500

QY 1501 KINSKDNHCNFMVLVNTQFCMALTDVT 1527
DB 1501 KINSKDNHCNFMVLVNTQFCMALTDVT 1527

RESULT 2
US-10-376-537-27
; Sequence 27, Application US/10376537
; Publication No. US20030224405A1
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042001
; CURRENT APPLICATION NUMBER: US/10/376,537
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US/09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 27
; LENGTH: 1527
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-376-537-27

Query Match      100.0%; Score 7967; DB 15; Length 1527;
```

Best Local Similarity 100.0%; Pred. No. 0; Matches 1527; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
Qy	1	MAPLLGRKPPPLVNPPLPGEEPPFTTIPHTQEAFTREEYEARELYSERIWTCKSTGSSQL	60
Db	1	MAPLLGRKPPPLVNPPLPGEEPPFTTIPHTQEAFTREEYEARELYSERIWTCKSTGSSQL	60
Qy	61	THKEAWEEQOEVAELLKEBFPWAEKLVLEMHVHNTASLEKLVDTAWLEIMTKYAVGEBEC	120
Db	61	THKEAWEEQOEVAELLKEBFPWAEKLVLEMHVHNTASLEKLVDTAWLEIMTKYAVGEBEC	120
Qy	121	DEEVGKEKMLVKVIKIHLEKVDDEATEKKSOGACDSSDKENSQIAQDHQKETYV	180
Db	121	DEEVGKEKMLVKVIKIHLEKVDDEATEKKSOGACDSSDKENSQIAQDHQKETYV	180
Qy	181	KEDEGRRESINDRARRSPKLPSTLKKGERKWAAPPKFLPHKYDVVKLQNEDKIISNPADS	240
Db	181	KEDEGRRESINDRARRSPKLPSTLKKGERKWAAPPKFLPHKYDVVKLQNEDKIISNPADS	240
Qy	241	LIRTERPPNKEIVRYFIRINNALRAGTGENAPWVEDELVKYKSLPSKFSDFLLDPYKYMT	300
Db	241	LIRTERPPNKEIVRYFIRINNALRAGTGENAPWVEDELVKYKSLPSKFSDFLLDPYKYMT	300
Qy	301	INPSTKRKNTGSPDRKPSKSKTDSNSSLSPINPLKWCVHLKKSLSGSPLVKYSKNSK	360
Db	301	INPSTKRKNTGSPDRKPSKSKTDSNSSLSPINPLKWCVHLKKSLSGSPLVKYSKNSK	360
Qy	361	SPEEHLEEMKMWSPNKLHNTFHI PKGPPAKPKGKHSKPLKAKGRSGILNGOKSTGN	420
Db	361	SPEEHLEEMKMWSPNKLHNTFHI PKGPPAKPKGKHSKPLKAKGRSGILNGOKSTGN	420
Qy	421	SKSPKKGLTKPTKMKQMTLLDMAKGTQKTRAPRNSGGTPTRTSSKPHKHLPPAALHLIA	480
Db	421	SKSPKKGLTKPTKMKQMTLLDMAKGTQKTRAPRNSGGTPTRTSSKPHKHLPPAALHLIA	480
Qy	481	YKENKORDEKDSALSCVLSKTARLLS SEDRALRPEELRSLVQKRYELLEHKRWASGE	540
Db	481	YKENKORDEKDSALSCVLSKTARLLS SEDRALRPEELRSLVQKRYELLEHKRWASGE	540
Qy	541	EQRKEYLKKRBEELKKLKEKAKERKEKEMLEKQKRYEQDELTKGNLPAPFLVDTP	600
Db	541	EQRKEYLKKRBEELKKLKEKAKERKEKEMLEKQKRYEQDELTKGNLPAPFLVDTP	600
Qy	601	GLPNTLFGDVAMVVEFLSCYSGILLPDAQYPI TAVSLMEALSADKGGFYLNARVLILQ	660
Db	601	GLPNTLFGDVAMVVEFLSCYSGILLPDAQYPI TAVSLMEALSADKGGFYLNARVLILQ	660
Qy	661	TLLODEIADYGEKLGKLSIEIPLTHSVSELVRLCLRRSDVQEESESGSDTDNKSAAPE	720
Db	661	TLLODEIADYGEKLGKLSIEIPLTHSVSELVRLCLRRSDVQEESESGSDTDNKSAAPE	720
Qy	721	DNEVDQEFLEKLETSFFELTSEKLIQILTALCHRLIMTYSVDHMETRQQMSAELWKER	780
Db	721	DNEVDQEFLEKLETSFFELTSEKLIQILTALCHRLIMTYSVDHMETRQQMSAELWKER	780
Qy	781	LAVLKEENDKRAEKQKREKEMAKNKENGKVENGLKTDKRIKRVKFEQVDTAEADMLS	840
Db	781	LAVLKEENDKRAEKQKREKEMAKNKENGKVENGLKTDKRIKRVKFEQVDTAEADMLS	840
Qy	841	AVKSRLLAIOAKKEREIOEREMKVKLERQABEERIRKHAAAEKAFQSGIAKAKLVMR	900
Db	841	AVKSRLLAIOAKKEREIOEREMKVKLERQABEERIRKHAAAEKAFQSGIAKAKLVMR	900
Qy	901	TPIGTDNRNRYLWFSDEVPGLPIBKGWVHDSIDYRFNHCKDHTVSGDEDCPRSKAN	960
Db	901	TPIGTDNRNRYLWFSDEVPGLPIBKGWVHDSIDYRFNHCKDHTVSGDEDCPRSKAN	960
Qy	961	LGNKASMTQHGTATAVEVETTPKQGNLWFLCDSQKELDELNCLHPQGIRESOLKER	1020
Db	961	LGNKASMTQHGTATAVEVETTPKQGNLWFLCDSQKELDELNCLHPQGIRESOLKER	1020
Qy	1021	LEKRYODIIHSHLARKPNLGLKSCDGNQELLNFLRSDLIEVATRLQKGLGVSETSEF	1080

Db	1021	LEKRYODIIHSHLARKPNLGLKSCDGNQELLNFLRSDLIEVATRLQKGLGVSETSEF	1080
Qy	1081	EARVISLEKLDKDFGECVIALQASVIKFKFLQGFMAPKQKRLQSEDSAKTEEVDESKQW	1140
Db	1081	EARVISLEKLDKDFGECVIALQASVIKFKFLQGFMAPKQKRLQSEDSAKTEEVDESKQW	1140
Qy	1141	BEAKVASALEKWKTAIREAQTFSRMHVLGLMDACIKWMSAENARCKVCPKKGEDDKLI	1200
Db	1141	BEAKVASALEKWKTAIREAQTFSRMHVLGLMDACIKWMSAENARCKVCPKKGEDDKLI	1200
Qy	1201	LCDECNKAPHLFCLPALYVEPDGEWQCPACOPATARRNSGRNNTYESASDESD	1260
Db	1201	LCDECNKAPHLFCLPALYVEPDGEWQCPACOPATARRNSGRNNTYESASDESD	1260
Qy	1261	ESEEEEEEEEDYEVAGLRPRKRTIRKHSVIPPAARSGRRPGCKPHSTRRSOPKAP	1320
Db	1261	ESEEEEEEEEDYEVAGLRPRKRTIRKHSVIPPAARSGRRPGCKPHSTRRSOPKAP	1320
Qy	1321	PVDDAEVDELVLQTKRSSRRQSLEKQCEIILHKIVKYRFSWPFPRPVTTRDEADYYDVI	1380
Db	1321	PVDDAEVDELVLQTKRSSRRQSLEKQCEIILHKIVKYRFSWPFPRPVTTRDEADYYDVI	1380
Qy	1381	THPMDPOTVONKSCSGSYRSVQEFLLTDMKQVFTNAEVYNCRGSHVLSCHVKTQCLVLL	1440
Db	1381	THPMDPOTVONKSCSGSYRSVQEFLLTDMKQVFTNAEVYNCRGSHVLSCHVKTQCLVLL	1440
Qy	1441	HGHLPGHPVRRKRKKFPDRLAEDEGDSEPAVAGSRDRDRSREAEIOEWLQDTSLYSA	1500
Db	1441	HGHLPGHPVRRKRKKFPDRLAEDEGDSEPAVAGSRDRDRSREAEIOEWLQDTSLYSA	1500
Qy	1501	KINSDHNCFMMLVNTQFCMALTDVT	1527
Db	1501	KINSDHNCFMMLVNTQFCMALTDVT	1527
RESULT 3			
US-10-702-148-27			
; Sequence 27, Application US/10702148			
; Publication No. US20040063145A1			
; GENERAL INFORMATION:			
; APPLICANT: Jones, Michael H.			
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR			
; FILE REFERENCE: 06501-042002			
; CURRENT APPLICATION NUMBER: US/10/702,148			
; CURRENT FILING DATE: 2003-11-05			
; PRIOR APPLICATION NUMBER: PRIOR APPLICATION NUMBER: US/09/839,479			
; PRIOR FILING DATE: 2001-04-20			
; PRIOR APPLICATION NUMBER: US 09/418,710.			
; PRIOR FILING DATE: 1999-10-15			
; PRIOR APPLICATION NUMBER: PCT/JP98/01783			
; PRIOR FILING DATE: 1998-04-17			
; PRIOR APPLICATION NUMBER: JP 9/310027			
; PRIOR FILING DATE: 1997-10-24			
; PRIOR APPLICATION NUMBER: JP 9/116570			
; PRIOR FILING DATE: 1997-04-18			
; NUMBER OF SEQ ID NOS: 72			
; SOFTWARE: FastSeq for Windows Version 4.0			
; SEQ ID NO 27			
; LENGTH: 1527			
; TYPE: PRT			
; ORGANISM: Homo sapiens			
US-10-702-148-27			
Query Match 100.0%; Score 1967; DB 15; Length 1527;			
Best Local Similarity 100.0%; Pred. No. 0;			
Matches 1527; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
Qy	1	MAPLLGRKPPPLVNPPLPGEEPPFTTIPHTQEAFTREEYEARELYSERIWTCKSTGSSOL	60
Db	1	MAPLLGRKPPPLVNPPLPGEEPPFTTIPHTQEAFTREEYEARELYSERIWTCKSTGSSOL	60
Qy	61	THKEAWEEQOEVAELLKEBFPWAEKLVLEMHVHNTASLEKLVDTAWLEIMTKYAVGEBEC	120

Db 61 THKEAWBEEQVAELLKBEFPAAWYBKLVLVEMVHNTASLEKLVDTAWLEIMTKYAVGSEC 120
QY 121 DFEVGEKXMLVKVIKIHIPLEKVDSEATEKSDGACDSPSSDKENSSQIAQDHQKKTIV 180
Db 121 DFEVGEKXMLVKVIKIHIPLEKVDSEATEKSDGACDSPSSDKENSSQIAQDHQKKTIV 180
QY 181 KEDEGRRESINDRARRSRKLPSTSLKGERKWAPKFLPHKYDYVKLQNEDKIISNVPADS 240
Db 181 KEDEGRRESINDRARRSRKLPSTSLKGERKWAPKFLPHKYDYVKLQNEDKIISNVPADS 240
QY 241 LIRTERPPNKEIVRYFIRHNALRACTGENAPWVVEDELVKKYSLSKFSDFLLDPYKMT 300
Db 241 LIRTERPPNKEIVRYFIRHNALRACTGENAPWVVEDELVKKYSLSKFSDFLLDPYKMT 300
QY 301 LNPSTKRKNTGSDPRKPKSKSTDNSSLSPLNPKLWCHVHLKLSLGSPLKXVNSKNSK 360
Db 301 LNPSTKRKNTGSDPRKPKSKSTDNSSLSPLNPKLWCHVHLKLSLGSPLKXVNSKNSK 360
QY 361 SPEEHLSEMMKMSPNKLTHTNPHIPKGGPPAKKPKGKSDKPLKAKGRSGILNGOKSTGN 420
Db 361 SPEEHLSEMMKMSPNKLTHTNPHIPKGGPPAKKPKGKSDKPLKAKGRSGILNGOKSTGN 420
QY 421 SKSPKKGKLTPTKTKMQMTLLDMAKGTQKMTAPRNSGGTPTSSKPKHKLPPAALHLIA 480
Db 421 SKSPKKGKLTPTKTKMQMTLLDMAKGTQKMTAPRNSGGTPTSSKPKHKLPPAALHLIA 480
QY 481 YKKNKREDKRSALSCVISTARLLSSDDRARLPEELRSVLQKRYELLEHKRWASMS 540
Db 481 YKKNKREDKRSALSCVISTARLLSSDDRARLPEELRSVLQKRYELLEHKRWASMS 540
QY 541 EQRKEVLLKKREBELKKLEKAKEREKEMLEKQRYEDELTKGNLPAFLVDTPE 600
Db 541 EQRKEVLLKKREBELKKLEKAKEREKEMLEKQRYEDELTKGNLPAFLVDTPE 600
QY 601 GLPNTLFGDVANVVFLLSCYSGULLPDAQYPTAVSLMEALSADKGGFLYLNRLVILLQ 660
Db 601 GLPNTLFGDVANVVFLLSCYSGULLPDAQYPTAVSLMEALSADKGGFLYLNRLVILLQ 660
QY 661 TLLQDEIAEDYGELGKMLSEIPLTHSVSELVRLCLRRSDVOESEGSDTDNKSAAFE 720
Db 661 TLLQDEIAEDYGELGKMLSEIPLTHSVSELVRLCLRRSDVOESEGSDTDNKSAAFE 720
QY 721 DNEVODEFLEKLETFEFPFELTSEKLOILTALCHRLMTYSQDHMETRQMSAELWKER 780
Db 721 DNEVODEFLEKLETFEFPFELTSEKLOILTALCHRLMTYSQDHMETRQMSAELWKER 780
QY 781 LAVLKEENDKKRAEKQKREMEAKNKENGKVENGLGKTDRKKRIVKFEPQVDTEADMIS 840
Db 781 LAVLKEENDKKRAEKQKREMEAKNKENGKVENGLGKTDRKKRIVKFEPQVDTEADMIS 840
QY 841 AVKSRRLLAIOAKGEREIOEREMVKLERQABEEIRIKHKAABEKAFOBGIKAKLVMR 900
Db 841 AVKSRRLLAIOAKGEREIOEREMVKLERQABEEIRIKHKAABEKAFOBGIKAKLVMR 900
QY 901 TPIGTDNRNHRWYLSDEVPGLFIEKGWVHDSIDYRFNHCHKDHVTSGDEDYCPRSKKN 960
Db 901 TPIGTDNRNHRWYLSDEVPGLFIEKGWVHDSIDYRFNHCHKDHVTSGDEDYCPRSKKN 960
QY 961 LGKNASMTQHGTATAEVAVETTTTQGGQNLWFLCDSQKDELLNCLHPQGIRSQLKER 1020
Db 961 LGKNASMTQHGTATAEVAVETTTTQGGQNLWFLCDSQKDELLNCLHPQGIRSQLKER 1020
QY 1021 LERKYODIITHSLARKPNGLKSCDGNQELLNFLRSDLIIEVATRLQKGLGYVETSEF 1080
Db 1021 LERKYODIITHSLARKPNGLKSCDGNQELLNFLRSDLIIEVATRLQKGLGYVETSEF 1080
QY 1081 EARVLSLEKLDKDFGECVIALQASVTKKFLQGGFMAPKQKRLQSDSASKEEVEDEKXW 1140
Db 1081 EARVLSLEKLDKDFGECVIALQASVTKKFLQGGFMAPKQKRLQSDSASKEEVEDEKXW 1140
QY 1141 EBAKVASALEKWKTAIREAQTFSRMHVLLGMLDACIKWDMSAENARCKVCPKKGDDKLI 1200
Db 1141 EBAKVASALEKWKTAIREAQTFSRMHVLLGMLDACIKWDMSAENARCKVCPKKGDDKLI 1200

QY 1201 LCDECNKAFHLFCRLPALYEPDGEWQCPACOPATARNRSRGRNTYTESASEDSBDDSD 1260
Db 1201 LCDECNKAFHLFCRLPALYEPDGEWQCPACOPATARNRSRGRNTYTESASEDSBDDSD 1260
QY 1261 EEEEEEEEEEEYEVAGLRRLPRKKTIRGKHSVIPPAARSGRRPGKKPHSTRSQKAP 1320
Db 1261 EEEEEEEEEEEYEVAGLRRLPRKKTIRGKHSVIPPAARSGRRPGKKPHSTRSQKAP 1320
QY 1321 PVDABVDELVLQTKRSRRQSLELQKCEEILHKIKVYKRFPSWPPREPVRDEAEYDVI 1380
Db 1321 PVDABVDELVLQTKRSRRQSLELQKCEEILHKIKVYKRFPSWPPREPVRDEAEYDVI 1380
QY 1381 THPMDFTQVQNKSCGYSYRQVEFLTDMKQVFTNAEVNCRGSHVLSWVWTEQCLVLL 1440
Db 1381 THPMDFTQVQNKSCGYSYRQVEFLTDMKQVFTNAEVNCRGSHVLSWVWTEQCLVLL 1440
QY 1441 HKHLPCHPYVRRKPKFPDRLAEDGDESEPAVGQSRDEDRRSREAEIQEWLQDTSLSA 1500
Db 1441 HKHLPCHPYVRRKPKFPDRLAEDGDESEPAVGQSRDEDRRSREAEIQEWLQDTSLSA 1500
QY 1501 KINSKDHNCFMMLVNTQFCMALTDVT 1527
Db 1501 KINSKDHNCFMMLVNTQFCMALTDVT 1527

RESULT 4

US-09-839-479-29
; Sequence 29, Application US/09839479
; Publication No. US20020039779A1
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042002
; CURRENT APPLICATION NUMBER: US/09/839,479
; CURRENT FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 29
; LENGTH: 1531
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-839-479-29

Query Match 99.8%; Score 7955; DB 9; Length 1531;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 1527; Conservative 0; Mismatches 0; Indels 4; Gaps 1;

QY 1 MAPLGRGPFPLVNPDPCEEPFPTIPHTQEAFTREEYEARLERYSERIWTCKTSSQL 60
Db 1 MAPLGRGPFPLVNPDPCEEPFPTIPHTQEAFTREEYEARLERYSERIWTCKTSSQL 60
QY 61 THKEAWBEEQVAELLKBEFPAAWYBKLVLVEMVHNTASLEKLVDTAWLEIMTKYAVGSEC 120
Db 61 THKEAWBEEQVAELLKBEFPAAWYBKLVLVEMVHNTASLEKLVDTAWLEIMTKYAVGSEC 120
QY 121 DFEVGEKXMLVKVIKIHIPLEKVDSEATEKSDGACDSPSSDKENSSQIAQDHQKKTIV 180
Db 121 DFEVGEKXMLVKVIKIHIPLEKVDSEATEKSDGACDSPSSDKENSSQIAQDHQKKTIV 180
QY 181 KEDEGRRESINDRARRSRKLPSTSLKGERKWAPKFLPHKYDYVKLQNEDKIISNVPADS 240
Db 181 KEDEGRRESINDRARRSRKLPSTSLKGERKWAPKFLPHKYDYVKLQNEDKIISNVPADS 240
QY 241 LIRTERPPNKEIVRYFIRHNALRACTGENAPWVVEDELVKKYSLSKFSDFLLDPYKMT 300

Db 241 LIRTERPPNKEIVRYIFIRHNALRAGTCENAPWVEDELVKYSLPSKFSDFLLDPKYMT 300
Qy 301 LNPSTKRKNTGSPDRKPSKSKTDNSLSPLNPKLWCHVHLKXSLSGSPKVKNSKNSK 360
Db 301 LNPSTKRKNTGSPDRKPSKSKTDNSLSLSPLNPKLWCHVHLKXSLSGSPKVKNSKNSK 360
Qy 361 SPEEHLEEMKMMSPNKLHTNFHI PKKGPAPKPGKSHSKPLKAKGRSGILNGQKSTGN 420
Db 361 SPEEHLEEMKMMSPNKLHTNFHI PKKGPAPKPGKSHSKPLKAKGRSGILNGQKSTGN 420
Qy 421 SKSPKGLKTPKTKMKOMTLDDMAKGTQKMTAPRNSGGTPTTSKPKHKLPPAALHLIA 480
Db 421 SKSPKGLKTPKTKMKOMTLDDMAKGTQKMTAPRNSGGTPTTSKPKHKLPPAALHLIA 480
Qy 481 YYKENKDREKRSALSCVISTARLLSSEDRARLPEELRSLVQRYELLEHKRWASMS 540
Db 481 YYKENKDREKRSALSCVISTARLLSSEDRARLPEELRSLVQRYELLEHKRWASMS 540
Qy 541 EQRKEYLKKREBELKKLKEKAKEREKEMLEKQRYEDQELTGKQLPAPFLVDTP 600
Db 541 EQRKEYLKKREBELKKLKEKAKEREKEMLEKQRYEDQELTGKQLPAPFLVDTP 600
Qy 601 GLPNTLFGDVAMVVEFLSCYSGLLIPDAQYPIITAVSLMEALSADKGGFLYLRVLI 657
Db 601 GLPNTLFGDVAMVVEFLSCYSGLLIPDAQYPIITAVSLMEALSADKGGFLYLRVLI 657
Qy 658 -LIQTLQDEIABDYGELGWLKSEIPLTLHSVSELVRLCLRRSDVQESSEGSTDDNKDS 716
Db 661 TLLQTLQDEIABDYGELGWLKSEIPLTLHSVSELVRLCLRRSDVQESSEGSTDDNKDS 720
Qy 717 AAPEDNEVDEFLKLETSSEFFELTSEELQILTALCHRLMTYSVDHMETRQMSAEL 776
Db 721 AAPEDNEVDEFLKLETSSEFFELTSEELQILTALCHRLMTYSVDHMETRQMSAEL 780
Qy 777 WKERLAVLKEENDKKAEBQKREMEAKNKENGKVENGLKTDKRRKIYKFEPOVDTEAB 836
Db 781 WKERLAVLKEENDKKAEBQKREMEAKNKENGKVENGLKTDKRRKIYKFEPOVDTEAB 840
Qy 837 DMISAVKSRLLAIQAKKREIQREREMVKLEBROAEEERIRKHAABKAFQBGIAKAL 896
Db 841 DMISAVKSRLLAIQAKKREIQREREMVKLEBROAEEERIRKHAABKAFQBGIAKAL 900
Qy 897 VMRRTPIGTRNHNRYWLSFDEVPGLFIEKGWVHDSIDYRFNHHCKDHTVSGDEDYCP 956
Db 901 VMRRTPIGTRNHNRYWLSFDEVPGLFIEKGWVHDSIDYRFNHHCKDHTVSGDEDYCP 960
Qy 957 KKANLGKNASMTQHGTAATEVAVETTPKQGQNLWFLCDSQKELDBLLNCLHPQGIRESQ 1016
Db 961 KKANLGKNASMTQHGTAATEVAVETTPKQGQNLWFLCDSQKELDBLLNCLHPQGIRESQ 1020
Qy 1017 LKERLEKRYQDIITHSTHLARKPNLGHKSCDGNQELLNFLRSDLIETVATRLQKGLGYVEE 1076
Db 1021 LKERLEKRYQDIITHSTHLARKPNLGHKSCDGNQELLNFLRSDLIETVATRLQKGLGYVEE 1080
Qy 1077 TSEFEARVISLEKLDKGECVIALQASVIKKFLQGFMAPKQRRKLQSDSAKTEEVDEE 1136
Db 1081 TSEFEARVISLEKLDKGECVIALQASVIKKFLQGFMAPKQRRKLQSDSAKTEEVDEE 1140
Qy 1137 KKWVEAKVASALEKWKTAIREAQTSRMHVLLGMLDACIKWDMSAENARCKVCPKKGED 1196
Db 1141 KKWVEAKVASALEKWKTAIREAQTSRMHVLLGMLDACIKWDMSAENARCKVCPKKGED 1200
Qy 1197 DKLILCECNKAPHLFCLRPALYEVDPGEWQCPACOPATARRNSRGRNVTESASDSDES 1256
Db 1201 DKLILCECNKAPHLFCLRPALYEVDPGEWQCPACOPATARRNSRGRNVTESASDSDES 1260
Qy 1257 DESDEEBEEEEEDYEVAGLRRLPRKTIIRGKHSVIPPAARSRRRPGKPHSTRSQ 1316
Db 1261 DESDEEBEEEEEDYEVAGLRRLPRKTIIRGKHSVIPPAARSRRRPGKPHSTRSQ 1320
Qy 1317 PKAPPVDDAEVDLVLQTKRSSRQSLQKCEBILHKKIVKRYFSPWPFREPVTRDEADY 1376

Db 1321 PKAPPVDDAEVDLVLQTKRSSRQSLQKCEBILHKKIVKRYFSPWPFREPVTRDEADY 1380
Qy 1377 YDVIITHPMDFOQTQVONKSCGSRYSVOEFLTDMKQVFTNAEVYNCRSHVLSCHVKTQCL 1436
Db 1381 YDVIITHPMDFOQTQVONKSCGSRYSVOEFLTDMKQVFTNAEVYNCRSHVLSCHVKTQCL 1440
Qy 1437 VLLHLKHLPGHPYVRRKRKKFPDRLAEBDEGDSPEAVGQSRDDEDRSRREABIOEWLQDTS 1496
Db 1441 VLLHLKHLPGHPYVRRKRKKFPDRLAEBDEGDSPEAVGQSRDDEDRSRREABIOEWLQDTS 1500
Qy 1497 LYSAKINSKDHNCFMMLVNTQFCMALTDTVT 1527
Db 1501 LYSAKINSKDHNCFMMLVNTQFCMALTDTVT 1531

RESULT 5

US-10-376-537-29
; Sequence 29, Application US/10376537
; Publication No. US20030224405A1
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042001
; CURRENT APPLICATION NUMBER: US/10/376,537
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US/09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 29
; LENGTH: 1531
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-376-537-29

Query Match 99.8%; Score 7955; DB 15; Length 1531;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 1527; Conservative 0; Mismatches 0; Indels 4; Gaps 1;

Qy 1 MAPLGRKFPPLVNPPLGSEPPFTTIPHTQEAFTREEYEARELRYSERIWTCKSTGSSQL 60
Db 1 MAPLGRKFPPLVNPPLGSEPPFTTIPHTQEAFTREEYEARELRYSERIWTCKSTGSSQL 60
Qy 61 THKEAWEESQEAELLKEEFPAYEKLVEWVHHNTASLEKLVDTAWLEIMTKYAVGEEC 120
Db 61 THKEAWEESQEAELLKEEFPAYEKLVEWVHHNTASLEKLVDTAWLEIMTKYAVGEEC 120
Qy 121 DFEVGKEMKLVKVIKIHPLKVDDEATEKSDGACDSSPSKSENSQIAQDHQKKEVTV 180
Db 121 DFEVGKEMKLVKVIKIHPLKVDDEATEKSDGACDSSPSKSENSQIAQDHQKKEVTV 180
Qy 181 KEDEGRRESINDRARRSPKLPSTSLKGERKWAAPPKFLPHKYDVYKLVQNEDKIISNVPADS 240
Db 181 KEDEGRRESINDRARRSPKLPSTSLKGERKWAAPPKFLPHKYDVYKLVQNEDKIISNVPADS 240
Qy 241 LIRTERPPNKEIVRYIFIRHNALRAGTCENAPWVEDELVKYSLPSKFSDFLLDPKYMT 300
Db 241 LIRTERPPNKEIVRYIFIRHNALRAGTCENAPWVEDELVKYSLPSKFSDFLLDPKYMT 300
Qy 301 LNPSTKRKNTGSPDRKPSKSKTDNSLSPLNPKLWCHVHLKXSLSGSPKVKNSKNSK 360
Db 301 LNPSTKRKNTGSPDRKPSKSKTDNSLSPLNPKLWCHVHLKXSLSGSPKVKNSKNSK 360
Qy 361 SPEEHLEEMKMMSPNKLHTNFHI PKKGPAPKPGKSHSKPLKAKGRSGILNGQKSTGN 420
Db 361 SPEEHLEEMKMMSPNKLHTNFHI PKKGPAPKPGKSHSKPLKAKGRSGILNGQKSTGN 420

QY 421 SKSPKGLKTPKTKMKTLLDMAKGTQKTRAPRNSGGTPTSSKPKHKLPPAALHLIA 480
DB 421 SKSPKGLKTPKTKMKTLLDMAKGTQKTRAPRNSGGTPTSSKPKHKLPPAALHLIA 480
QY 481 YYKENKREDKRSALSVCISKTARLLSSEDRARLPEELRSIVQKRYELLEHKRWASME 540
DB 481 YYKENKREDKRSALSVCISKTARLLSSEDRARLPEELRSIVQKRYELLEHKRWASME 540
QY 541 EQRKEYLKKREELKKLKEKAKERREKEMLERLEKQKRYEDQELTGKMLPAFRLVDTPE 600
DB 541 EQRKEYLKKREELKKLKEKAKERREKEMLERLEKQKRYEDQELTGKMLPAFRLVDTPE 600
QY 601 GLPNTLFGDVAWVEFLSCYSGLLLPDAQPYPTAVSLMEALSADGGFLYLNRLVILQ 657
DB 601 GLPNTLFGDVAWVEFLSCYSGLLLPDAQPYPTAVSLMEALSADGGFLYLNRLVILQ 660
QY 658 -LLQTLLODETAEDYGELGMKLSPLTLHSVSELVRLCLRRSDVQEESESGSDTDNDKDS 716
DB 661 TLLOTLLODETAEDYGELGMKLSPLTLHSVSELVRLCLRRSDVQEESESGSDTDNDKDS 720
QY 717 AAFEDNEVQDEFLEKLTSEPFELTSEBKLIITLALCHRIILMTYSVQDHMETROQMSAEL 776
DB 721 AAFEDNEVQDEFLEKLTSEPFELTSEBKLIITLALCHRIILMTYSVQDHMETROQMSAEL 780
QY 777 WKERLAVLKEENDKRAEKQKKEAKENKENGKVENGLGKTDRKRIVKPEPQVDTAE 836
DB 781 WKERLAVLKEENDKRAEKQKKEAKENKENGKVENGLGKTDRKRIVKPEPQVDTAE 840
QY 837 DMISAVKRRLLAIQAKKEREIOEREMKVQLERQAEERIRKHAAAEKAFQEGIAKAKL 896
DB 841 DMISAVKRRLLAIQAKKEREIOEREMKVQLERQAEERIRKHAAAEKAFQEGIAKAKL 900
QY 897 VMRRTPIGTRNHNRYLWFSDEVPLGFIKGGVHDSIDYRFNHHCQDHTVSGDEYDCPRS 956
DB 901 VMRRTPIGTRNHNRYLWFSDEVPLGFIKGGVHDSIDYRFNHHCQDHTVSGDEYDCPRS 960
QY 957 KXANTGKXASMTQGTATEVAVETTPKQGNLWFLCDSQKELDELNCLHPOGIRSEQ 1016
DB 961 KXANTGKXASMTQGTATEVAVETTPKQGNLWFLCDSQKELDELNCLHPOGIRSEQ 1020
QY 1017 LKERLEKYQDIITHIHLARKPNLGLKSCDGNQELLNFRSLDIEVATRLQKGLGYVEE 1076
DB 1021 LKERLEKYQDIITHIHLARKPNLGLKSCDGNQELLNFRSLDIEVATRLQKGLGYVEE 1080
QY 1077 TSEFARVISLEKLDQFCEVIALQASVIKKFLOGFMAPKQKRRKLQSEDSAKTEEVDEE 1136
DB 1081 TSEFARVISLEKLDQFCEVIALQASVIKKFLOGFMAPKQKRRKLQSEDSAKTEEVDEE 1140
QY 1137 KQWVEAKVASALEKWKTAIREAQTFSRMHVLLGMLDACIKWMSAENARCKVCPKKGED 1196
DB 1141 KQWVEAKVASALEKWKTAIREAQTFSRMHVLLGMLDACIKWMSAENARCKVCPKKGED 1200
QY 1197 DKLLILCDECNKAFHLCFLRPALYVDPGWCQCPACQATARNRSGRNYTSEASDESSED 1256
DB 1201 DKLLILCDECNKAFHLCFLRPALYVDPGWCQCPACQATARNRSGRNYTSEASDESSED 1260
QY 1257 DESDEEEEEEEDYEVAGLRLPRKTIIRKHSVIPPAARSGRPGKXPHSTRSQ 1316
DB 1261 DESDEEEEEEEDYEVAGLRLPRKTIIRKHSVIPPAARSGRPGKXPHSTRSQ 1320
QY 1317 PKAPPVDDAEVDDELVLQTKRSSRROSLEQKCEETLHKIKVKYRFSWPFREPTVTRAEADY 1376
DB 1321 PKAPPVDDAEVDDELVLQTKRSSRROSLEQKCEETLHKIKVKYRFSWPFREPTVTRAEADY 1380
QY 1377 YDVI THPMDFTQVQKSCGYSRYSVQEFPLTDMKQVFTNAEVNCRGSHVLSQWVTEOCL 1436
DB 1381 YDVI THPMDFTQVQKSCGYSRYSVQEFPLTDMKQVFTNAEVNCRGSHVLSQWVTEOCL 1440
QY 1437 VVLLHKHLPCHPYVRRKRRKKFPDRLAEDGDSPEAVQSGRDEDRRSREAEIOEWLQDTS 1496
DB 1441 VVLLHKHLPCHPYVRRKRRKKFPDRLAEDGDSPEAVQSGRDEDRRSREAEIOEWLQDTS 1500
QY 1497 LYSAKINSKDHNCFMMLVNTQFCMALTDVT 1527

DB 1501 LYSAKINSKDHNCFMMLVNTQFCMALTDVT 1531
RESULT 6
US-10-702-148-29
; Sequence 29, Application US/10702148
; Publication No. US20040063145A1
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042002
; CURRENT APPLICATION NUMBER: US/10/702,148
; CURRENT FILING DATE: 2003-11-05
; PRIOR APPLICATION NUMBER: PRIOR APPLICATION NUMBER: US/09/839,479
; PRIOR FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 29
; LENGTH: 1531
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-702-148-29

Query Match 99.8%; Score 7955; DB 15; Length 1531;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 1527; Conservative 0; Mismatches 0; Indels 4; Gaps 1;

QY 1 MAPLGRPFPLVNPPLPCGEEFFFTIPHTQEAFTREEYEARLERYSERIWTCKSGSSOL 60
DB 1 MAPLGRPFPLVNPPLPGEEFFFTIPHTQEAFTREEYEARLERYSERIWTCKSGSSOL 60
QY 61 THKEAWEEQVEAEELLKEEFPANWEKLVLEMVHHTASLEKLVDTAWLEIMTKYAVGESC 120
DB 61 THKEAWEEQVEAEELLKEEFPANWEKLVLEMVHHTASLEKLVDTAWLEIMTKYAVGESC 120
QY 121 DFEVGKEXMLKVKIWIHPLEKVDDEATEKSDGACDSPSSDKENSSQIAQDHQKKTVV 180
DB 121 DFEVGKEXMLKVKIWIHPLEKVDDEATEKSDGACDSPSSDKENSSQIAQDHQKKTVV 180
QY 181 KEDEGRRESINDRARRSRPKLPTSLKGERKWAAPKFLPHKYVDVKLQNEDKIIISNPADS 240
DB 181 KEDEGRRESINDRARRSRPKLPTSLKGERKWAAPKFLPHKYVDVKLQNEDKIIISNPADS 240
QY 241 LIRTERPNKEIVRYFIRHNALRAGTGENAPWVVEDELVKKYSLPKSFDFLLDPYKYWT 300
DB 241 LIRTERPNKEIVRYFIRHNALRAGTGENAPWVVEDELVKKYSLPKSFDFLLDPYKYWT 300
QY 301 LNPSTKRKNTGSPDRPKSKKTDNSSLSSPLNPKLWCHVHLKKSLSGSPLVKXNSKNSK 360
DB 301 LNPSTKRKNTGSPDRPKSKKTDNSSLSSPLNPKLWCHVHLKKSLSGSPLVKXNSKNSK 360
QY 361 SPEEHLEEMKXNSKNTLHTNPHIPKGGPPAKKPKGKSDKPLKAKGRSKGIINGOKSTGN 420
DB 361 SPEEHLEEMKXNSKNTLHTNPHIPKGGPPAKKPKGKSDKPLKAKGRSKGIINGOKSTGN 420
QY 421 SKSPKGLKTPKTKMKTLLDMAKGTQKTRAPRNSGGTPTSSKPKHKLPPAALHLIA 480
DB 421 SKSPKGLKTPKTKMKTLLDMAKGTQKTRAPRNSGGTPTSSKPKHKLPPAALHLIA 480
QY 481 YYKENKREDKRSALSVCISKTARLLSSEDRARLPEELRSIVQKRYELLEHKRWASME 540
DB 481 YYKENKREDKRSALSVCISKTARLLSSEDRARLPEELRSIVQKRYELLEHKRWASME 540
QY 541 EQRKEYLKKREELKKLKEKAKERREKEMLERLEKQKRYEDQELTGKMLPAFRLVDTPE 600

```
Db 541 EQRKEYLKKRBEELKKLKEKAKERKEKLEKQVEDQELTGKNLPAPRLVDTP 600
Qy 601 GLPNTLFGDVAMVVEFLSCYSGLLLPDAQYPTAVSLMEALSADKGGFLYLNRLVI 657
Db 601 GLPNTLFGDVAMVVEFLSCYSGLLLPDAQYPTAVSLMEALSADKGGFLYLNRLVILQ 660
Qy 658 -LLQTLQDIBIADYGBELGKMLSEIPLTHSVSELVRLCLRRSDVQEESEGSDTDNKKOS 716
Db 661 TLLQTLQDIBIADYGBELGKMLSEIPLTHSVSELVRLCLRRSDVQEESEGSDTDNKKOS 720
Qy 717 AAFEDNEVDPELEKLETSFPPELTSEKLOILTALCHRLMTYSVODHMETROQMSAEL 776
Db 721 AAFEDNEVDPELEKLETSFPPELTSEKLOILTALCHRLMTYSVODHMETROQMSAEL 780
Qy 777 WKERLAVLEENDKGAEKQKREMEAKNKENGVENGLKTDKRRKRIVKFEPQVDTEAE 836
Db 781 WKERLAVLEENDKGAEKQKREMEAKNKENGVENGLKTDKRRKRIVKFEPQVDTEAE 840
Qy 837 DMISAVKSRLLAIQAKKEREIOEREMKVLERQAEERIRKHAATAKAFQBGIAKAKL 896
Db 841 DMISAVKSRLLAIQAKKEREIOEREMKVLERQAEERIRKHAATAKAFQBGIAKAKL 900
Qy 897 VMERTPIGTDRNHRWYLFSDVEPGLFIEKGWYHDSIDYRFNHHCKDHTVSGDEDYCPRS 956
Db 901 VMERTPIGTDRNHRWYLFSDVEPGLFIEKGWYHDSIDYRFNHHCKDHTVSGDEDYCPRS 960
Qy 957 KKANLGKNSAMTQHGTA TEVA VETTPKQGNLWFLCDSOKELDELLNCLHPQGIRESQ 1016
Db 961 KKANLGKNSAMTQHGTA TEVA VETTPKQGNLWFLCDSOKELDELLNCLHPQGIRESQ 1020
Qy 1017 LKERLEKRYQDI IHSIHLARKPNLGLKSCDGNQELNLFRLSDLIEVATRLQKGLGYVEE 1076
Db 1021 LKERLEKRYQDI IHSIHLARKPNLGLKSCDGNQELNLFRLSDLIEVATRLQKGLGYVEE 1080
Qy 1077 TSEFEARVTSLEKLDKDFGECVIALQASVKKFLOGFMAPKQKRKLQSDS AKTESEVDEE 1136
Db 1081 TSEFEARVTSLEKLDKDFGECVIALQASVKKFLOGFMAPKQKRKLQSDS AKTESEVDEE 1140
Qy 1137 KQWVEAKVASALEKWKTAIREAQTFSRMHVLGLMDACIKWDMSAENARCKVCPKKGED 1196
Db 1141 KQWVEAKVASALEKWKTAIREAQTFSRMHVLGLMDACIKWDMSAENARCKVCPKKGED 1200
Qy 1197 DKLILDECNKAFHLCLRPALYVEPDGEMQCPACQATARRNSRGRNTYTESASDSED 1256
Db 1201 DKLILDECNKAFHLCLRPALYVEPDGEMQCPACQATARRNSRGRNTYTESASDSED 1260
Qy 1257 DESDEEBEEEEEDYEVAGLRPRKTI RGHKSVIPPAARSORRRPKKPHSTRRSQ 1316
Db 1261 DESDEEBEEEEEDYEVAGLRPRKTI RGHKSVIPPAARSORRRPKKPHSTRRSQ 1320
Qy 1317 PKAPPVDDAEVDLVLQTKRSSRQSLQKCEIILHKIVKRFSPRPREPVRTRDAEDY 1376
Db 1321 PKAPPVDDAEVDLVLQTKRSSRQSLQKCEIILHKIVKRFSPRPREPVRTRDAEDY 1380
Qy 1377 YDVITHPMDFQTVQNKSCGYSRVQEFJLTDKMQVFTNAEVYNCRSHVLS CMVKTEOCL 1436
Db 1381 YDVITHPMDFQTVQNKSCGYSRVQEFJLTDKMQVFTNAEVYNCRSHVLS CMVKTEOCL 1440
Qy 1437 VLLHKLHCPHVPVRRKPKFPDRLAEDGDSPEAVGQSRDDEDRSRBAETQEWLQDTS 1496
Db 1441 VLLHKLHCPHVPVRRKPKFPDRLAEDGDSPEAVGQSRDDEDRSRBAETQEWLQDTS 1500
Qy 1497 LYSAKINSKDHNCFMMLVNTQFCMALTDTVT 1527
Db 1501 LYSAKINSKDHNCFMMLVNTQFCMALTDTVT 1531
```

RESULT 7

US-09-839-479-68

; Sequence 68, Application US/09839479

; Publication No. US20020039779A1

; GENERAL INFORMATION:

```
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042002
; CURRENT APPLICATION NUMBER: US/09/839,479
; CURRENT FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 68
; LENGTH: 1525
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-839-479-68
```

```
Query Match 99.0%; Score 7888; DB 9; Length 1525;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 1521; Conservative 1; Mismatches 2; Indels 4; Gaps 4;
```

```
Qy 1 MAPLLGRKFPFLVNPPLPGSEPFPTTIPHTQEAFTREYEREARLERYSERIWTCKTGSSQL 60
Db 1 MAPLLGRKFPFLVNPPLPGSEPFPTTIPHTQEAFTREYEREARLERYSERIWTCKTGSSQL 60
Qy 61 THKEAWEERQEVAAELKEEFPAWYEKLVLEMHHTNTASLEKLVDTAWLEIMTYKAVGEEC 120
Db 61 THKEAWEERQEVAAELKEEFPAWYEKLVLEMHHTNTASLEKLVDTAWLEIMTYKAVGEEC 120
Qy 121 DPEVGKEMLVKVIKVIHLEKVDDEATEKKSDDGACDSSPSSDKENSSQIAQDHOKKETT 180
Db 121 DPEVGKEMLVKVIKVIHLEKVDDEATEKKSDDGACDSSPSSDKENSSQIAQDHOKKETT 180
Qy 181 KEDEGRRESINDRARRSPRKLPTSLKKGKRWAPPKFLPHKYDVKLQNEDKIISNYPADS 240
Db 181 KEDEGRRESINDRARRSPRKLPTSLKKGKRWAPPKFLPHKYDVKLQNEDKIISNYPADS 240
Qy 241 LIRTERPPNKEIVRYFIRINLRAGTGENAPWVVEDELVKYKSLPSKFSDFLLDPKYMT 300
Db 241 LIRTERPPNKEIVRYFIRINLRAGTGENAPWVVEDELVKYKSLPSKFSDFLLDPKYMT 300
Qy 301 LNPSTKRKNTGSPDRKPSKSKTDSNLSSPLNPKLWCHVHLKKSLSGSPKVKNSKNSK 360
Db 301 LNPSTKRKNTGSPDRKPSKSKTDSNLSSPLNPKLWCHVHLKKSLSGSPKVKNSKNSK 360
Qy 361 SPEEHLEENMKWMSPNKLTNPHIPKKGPPAKKPGKHSKPLKAKGRSGILNGQKSTGN 420
Db 361 SPEEHLEENMKWMSPNKLTNPHIPKKGPPAKKPGKHSKPLKAKGRSGILNGQKSTGN 420
Qy 421 SKSPKKGLTKPTKMKQMTLLDMAKGTQKWTAPRNSGGTPTRTSSKPKHKLHPAALHLIA 480
Db 421 SKSPKKGLTKPTKMKQMTLLDMAKGTQKWTAPRNSGGTPTRTSSKPKHKLHPAALHLIA 480
Qy 481 YYKENKORDEKESALS CVLSKTARLLSSEDRARLPEELSLVQKRYELLEHKRWASMS 540
Db 481 YYKENKORDEKESALS CVLSKTARLLSSEDRARLPEELSLVQKRYELLEHKRWASMS 540
Qy 541 EQRKEYLKKRBEELKKLKEKAKERKEKEMLEKQRYEDELQELTGKNLPAPRLVDTP 600
Db 541 EQRKEYLKKRBEELKKLKEKAKERKEKEMLEKQRYEDELQELTGKNLPAPRLVDTP 600
Qy 601 GLPNTLFGDVAMVVEFLSCYSGLLLPDAQYPTAVSLMEALSADKGGFLYLNRLVILQ 660
Db 601 GLPNTLFGDVAMVVEFLSCYSGLLLPDAQYPTAVSLMEALSADKGGFLYLNRLVILQ 660
Qy 661 TLLQDIBIADYGBELGKMLSEIPLTHSVSELVRLCLRRSDVQEESEGSDTDNKKOSAAFE 720
Db 660 TLLQDIBIADYGBELGKMLSEIPLTHSVSELVRLCLRRSDVQEESEGSDTDNKKOSAAFE 719
```


901 TP1GTDNRNHRNYLWFSDEVPLGFIKGGWVHDSIDYRPNHCKDHTVSGDEDCYPRSKKAN 960
900 TP1GTDNRNHRNYLWFSDEVPLGFIKGGWVHDSIDYRPNHCKDHTVSGDEDCYPRSKKAN 959
961 LKGNASMTQHGCTATEVAVETTTTPKOGONLWFLCDSQKELDELLNCLHPQIGRESOLKER 1020
960 LKGNASMTQHGCTATEVAVETTTTPKOGONLWFLCDSQKELDELLNCLHPQIGRESOLKER 1019
1021 LEKRYQDI1HSIHLAKRPNLGLKSCDGNQELNLFRLSDLIETVATRLQKGLGVVETSEF 1080
1020 LEKRYQDI1HSIHLAKRPNLGLKSCDGNQELNLFRLSDLIETVATRLQKGLGVVETSEF 1079
1081 EARVISLEKUKDFGECVIALQASVIKKFLQGFMAPKQRRKLOESDAKTEEVDEEKKWV 1140
1080 EARVISLEKUKDFGECVIALQASVIKKFLQGFMAPKQRRKLOESDAKTEEVDEEKKWV 1139
1141 EEAIVASALEKWKTAIREAQTFSRMHVLLGMLDACIKWMSAENARCKVCPKKGEDDKLI 1200
1140 EEAIVASALEKWKTAIREAQTFSRMHVLLGMLDACIKWMSAENARCKVCPKKGEDDKLI 1199
1201 LCDECNKAFHLCLRPALVEVPDGEWQCPACQATARRNSRGRNYYTEESASEDESD 1260
1200 LCDECNKAFHLCLRPALVEVPDGEWQCPACQATARRNSRGRNYYTEESASEDESD 1258
1261 EEEEEEEEEEDYEVAGRLRPRKTIIRGKHSVIPPAARSGRRPGKPHSTRSQPKAP 1320
1259 EEEEEEEEEEDYEVAGRLRPRKTIIRGKHSVIPPAARSGRRPGKPHSTRSQPKAP 1318
1321 PVDVDAEVLVLQTKSSRRQSLQKCEBILHKIVKIRFSPWFRPVRDEADYYDVI 1380
1319 PV-DAEVLVLQTKSSRRQSLQKCEBILHKIVKIRFSPWFRPVRDEADYYDVI 1377
1381 THPMDFTQVQKCSGYSRSVQBFLLTDMKQVFNAAEYVNCRGSHVLSCHVKTBOCLVLL 1440
1378 THPMDFTQVQKCSGYSRSVQBFLLTDMKQVFNAAEYVNCRGSHVLSCHVKTBOCLVLL 1437
1441 HKHLPCHPYVRRKRRKFPDLAEDGSEPEAVGQSRDEDRRSREAEIOEWLQDTSLY-S 1499
1438 HKHLPCHPYVRRKRRKFPDLAEDGSEPEAVGQSRDEDRRSREAEIOEWLQDTSLYAS 1497
1500 AKINSKDHNCFMMLVNTQFCMALTDVT 1527
1498 AKINSKDHNCFMMLVNTQFCMALTDVT 1525

RESULT 9

US-10-702-148-68
; Sequence 68, Application US/10702148
; Publication No. US20040063145A1
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042002
; CURRENT APPLICATION NUMBER: US/10/702,148
; CURRENT FILING DATE: 2003-11-05
; PRIOR APPLICATION NUMBER: PRIOR APPLICATION NUMBER: US/09/839,479
; PRIOR FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 68
; LENGTH: 1525
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-702-148-68

Query Match 99.0%; Score 7888; DB 15; Length 1525;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 1521; Conservative 1; Mismatches 2; Indels 4; Gaps 4;
QY 1 MAPLLGRKPPPLVNPPLGSEPPFTI PHTQEAFTREBYEARLERYERIWTCKSTGSSQL 60
DB 1 MAPLLGRKPPPLVNPPLGSEPPFTI PHTQEAFTREBYEARLERYERIWTCKSTGSSQL 60
QY 61 THKEAWEEOEVAELKEBFPWYKLVLEMHVHNTASLEKVDTAWELEMTKYAVGEEC 120
DB 61 THKEAWEEOEVAELKEBFPWYKLVLEMHVHNTASLEKVDTAWELEMTKYAVGEEC 120
QY 121 DFEVGEKMLKVKIHKHLEKVEDEATEKSDGACDSSPSDKENSQIAQDHQKKEVTV 180
DB 121 DFEVGEKMLKVKIHKHLEKVEDEATEKSDGACDSSPSDKENSQIAQDHQKKEVTV 180
QY 181 KEDEGRRESINDRARRSPRKLPTSLKKGKRWAPPKFLPHKYDVKLQNEDKIISNVPADS 240
DB 181 KEDEGRRESINDRARRSPRKLPTSLKKGKRWAPPKFLPHKYDVKLQNEDKIISNVPADS 240
QY 241 LIRTERPPNKEIVRYFIRHNALRAGTGENAPWVEDELVKYSLPSKFSDFLLDPYKMT 300
DB 241 LIRTERPPNKEIVRYFIRHNALRAGTGENAPWVEDELVKYSLPSKFSDFLLDPYKMT 300
QY 301 LNPSTKRKMTGSPDRKPSKSKTDNSSLSPLNPKLWCHVHLKKSLSGSPKVKNSKNSK 360
DB 301 LNPSTKRKMTGSPDRKPSKSKTDNSSLSPLNPKLWCHVHLKKSLSGSPKVKNSKNSK 360
QY 361 SPEEHLLEEMKMSPNKLTNPHI PKGPPAKPGKHSDKPLKAKGRSGILNGQKSTGN 420
DB 361 SPEEHLLEEMKMSPNKLTNPHI PKGPPAKPGKHSDKPLKAKGRSGILNGQKSTGN 420
QY 421 SKSPKGLKTPKTKMQMTLLDMKGTQKWTAPRNSGGTPTRTSKPHKHLPPAALHLIA 480
DB 421 SKSPKGLKTPKTKMQMTLLDMKGTQKWTAPRNSGGTPTRTSKPHKHLPPAALHLIA 480
QY 481 YYKENKDRDEKRSALSCVISTARLISSEDRARLPEELRSLVOKRYELLEHKRWASMS 540
DB 481 YYKENKDRDEKRSALSCVISTARLISSEDRARLPEELRSLVOKRYELLEHKRWASMS 540
QY 541 EQRKEYLKKRREBELKKLKEKAKEREKEMLEKQRYEDQELTGKMLPAFRLVDTPE 600
DB 541 EQRKEYLKKRREBELKKLKEKAKEREKEMLEKQRYEDQELTGKMLPAFRLVDTPE 600
QY 601 GLPNTLFGDVAMVVFELSCVSGLLIPDAQYPTAVSLMEALSADKGGFLYLRVLLVLLQ 660
DB 601 GLPNTLFGDVAMVVFELSCVSGLLIPDAQYPTAVSLMEAL-ADKGGFLYLRVLLVLLQ 659
QY 661 TLLQDEIAEDYGELGMLKSEI PLTLHSVSELVRLCLRRSDVOEESGSDTDDNKDSAAFE 720
DB 661 TLLQDEIAEDYGELGMLKSEI PLTLHSVSELVRLCLRRSDVOEESGSDTDDNKDSAAFE 719
QY 721 DNEVQDEFLEKLETSFEFFELTSEELQILTALCHRIILMTYSVQDHMETRQMSAELWKR 780
DB 721 DNEVQDEFLEKLETSFEFFELTSEELQILTALCHRIILMTYSVQDHMETRQMSAELWKR 779
QY 781 LAVLEENDKKAEBKQKREKEMAKENKGVNGLKTDKRIIVKFEPOVDTEAEDMTS 840
DB 781 LAVLEENDKKAEBKQKREKEMAKENKGVNGLKTDKRIIVKFEPOVDTEAEDMTS 839
QY 841 AVKSRRLLAIOAKKEREIOEREMKVKLERQAEERIRKHAAAEKAFQGGIAKAKLVMMR 900
DB 841 AVKSRRLLAIOAKKEREIOEREMKVKLERQAEERIRKHAAAEKAFQGGIAKAKLVMMR 899
QY 901 TPIGTDNRNHRNYLWFSDEVPLGFIKGGWVHDSIDYRPNHCKDHTVSGDEDCYPRSKKAN 960
DB 901 TPIGTDNRNHRNYLWFSDEVPLGFIKGGWVHDSIDYRPNHCKDHTVSGDEDCYPRSKKAN 959
QY 961 LKGNASMTQHGCTATEVAVETTTTPKOGONLWFLCDSQKELDELLNCLHPQIGRESOLKER 1020
DB 961 LKGNASMTQHGCTATEVAVETTTTPKOGONLWFLCDSQKELDELLNCLHPQIGRESOLKER 1019

Db 481 KERLAVLKEENDKRAEKQKREMEAKNKGKNGVGLGKTRKKEIVKFPQVDTAEAD 540
Qy 838 MISAVKSRELLAIQAKKEREIOEREMKVKL 867
Db 541 MISAVKSRELLAIQAKKEREIOEREMKVKI 570

RESULT 11
US-09-864-761-37938
; Sequence 37938, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecm1ca-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 37938
; LENGTH: 560
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC005074.1
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.5
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2.8
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.9
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.93
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.99

; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.8
; OTHER INFORMATION: EST HUMAN HIT: AW976211.1, EVALUATE 2.00e-91
; OTHER INFORMATION: SWISSPROT HIT: P53935, EVALUATE 2.00e-03
US-09-864-761-37938

Query Match 35.7%; Score 2847; DB 9; Length 560;
Best Local Similarity 99.8%; Pred. No. 9.1e-153;
Matches 559; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 305 TTRKNTGSPDRKPSKSKTDNSLSPLNPKLVCHVHLKKSLSGSLKVKNSKNSKSP 364
Db 1 TTRKNTGSPDRKPSKSKTDNSLSPLNPKLVCHVHLKKSLSGSLKVKNSKNSKSP 60
Qy 365 HLEENMKWMSPNKLTNFIHFKGPPAKPGKHSDKPLKAKGSKGILNGKSTGNSKSP 424
Db 61 HLEENMKWMSPNKLTNFIHFKGPPAKPGKHSDKPLKAKGSKGILNGKSTGNSKSP 120
Qy 425 KKGLKTPKTKMOMTLDDWAKGTQKMTAPRNSGGTPTSSKPHKHLPPAALHLIAYYKE 484
Db 121 KKGLKTPKTKMOMTLDDWAKGTQKMTAPRNSGGTPTSSKPHKHLPPAALHLIAYYKE 180
Qy 485 NKDREDKRSALSCVISKTARLLSSDDRARLPBELRSLVQRYELLEHKKRWASMSBEQRK 544
Db 181 NKDREDKRSALSCVISKTARLLSSDDRARLPBELRSLVQRYELLEHKKRWASMSBEQRK 240
Qy 545 EYLKKRBEELKKLKEKAKEREKEMLEKQRYEDQELTGKQLPAPFLVDTPEGLPN 604
Db 241 EYLKKRBEELKKLKEKAKEREKEMLEKQRYEDQELTGKQLPAPFLVDTPEGLPN 300
Qy 605 TLFQDVAMVVEFLSCYSGLLLPDAQYPTAVSLMEALSADKGGFLYLNRLVILLQTLQ 664
Db 301 TLFQDVAMVVEFLSCYSGLLLPDAQYPTAVSLMEALSADKGGFLYLNRLVILLQTLQ 360
Qy 665 DEIAEDYGEELGKMLSEIPLTHSVSELVRLCLRRSDVQSESGSDTDDNKDAAAFEDNEV 724
Db 361 DEIAEDYGEELGKMLSEIPLTHSVSELVRLCLRRSDVQSESGSDTDDNKDAAAFEDNEV 420
Qy 725 QDEFLEKLETSEFFELTSEEKLIQTLALCHRLIMTYSVQDHMETROQMAELWKERLAVL 784
Db 421 QDEFLEKLETSEFFELTSEEKLIQTLALCHRLIMTYSVQDHMETROQMAELWKERLAVL 480
Qy 785 KEENDKRAEKQKREMEAKNKGKNGVGLGKTRKKEIVKFPQVDTAEADMSAVKS 844
Db 481 KEENDKRAEKQKREMEAKNKGKNGVGLGKTRKKEIVKFPQVDTAEADMSAVKS 540
Qy 845 RLLAIQAKKEREIOEREMK 864
Db 541 RLLAIQAKKEREIOEREMK 560

RESULT 12
US-09-839-479-1
; Sequence 1, Application US/09839479
; Publication No. US20020039779A1
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042002
; CURRENT APPLICATION NUMBER: US/09/839,479
; CURRENT FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1674

```
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-839-479-1

Query Match      11.7%; Score 929.5; DB 9; Length 1674;
Best Local Similarity 23.1%; Pred. No. 9.4e-44;
Matches 395; Conservative 264; Mismatches 529; Indels 525; Gaps 67;

QY 3 PLLGKPP-----PLVNPILGPEPFTTIPHTQAFRTREYEAERLRYERIWTCKSTGSS 58
Db ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
152 PLLHKKPFVRQKPPADLAPDEEVFY-CKVTNEIFRHYDDFFERTILCNLSVMSCAVTGRP 210
QY 59 QLTKEAMEEEQVALLKEBPANVEKLVLEMVH-----HNTASLEKLVDTAWLEIMTKY 114
Db ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
211 GLTYQEALESEKAKQNLQ--QSPF---EPLIIPVLYLTSLTHRSRLHEICDDIFAVVKORY 266
QY 115 AVGECDDPEVGEKMKLVKVIKIHPLEKVEDATEKKGSDGACDSPSSDKENSSQIAQDHQ 174
Db ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
267 FVEETVEVRNNGARLOCTILEVLP-----PS-----HQNGFANGHV 303
QY 175 KK---ETVVKDEGRRESINDARRSPKPLPTSLKKGKRWKAPPKPLPHKYDV-----KLQ 227
Db ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
304 NSVDGETIISDSDSETQS-----CSFQNGKKKDAIDPLL-FKYKVQPTKKEL 351
QY 228 NEDKIISNPADSLIRTPPNKEIVRFIRHNALRAGTGENAPMVVEDELVKKYSLPS- 286
Db ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
352 HESAI---VKATQISRRKHLFRDKLULFQHC-----EPQGVIK---IFASSLSY 399
QY 287 -----KFSDFLDLPKYMTLPNSTKRNKNTGSPDRKPSK---SKTDNSLSLSPNKLWC 338
Db ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
400 KIAQDFSVFFDDPTFTFSPANRRG-----RPPKRIHISOEDN----- 440
QY 339 VHYLKKSLSGSLPKVNSKNKSPBEHLEEMMKOMSPNKLHNFHIPPKGPPAKPKGKHS 398
Db ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
441 -VANKOTLA-----SYRSKATKER----- 458
QY 399 DKPLAKAGSKILNGQKSTGNSKSPKGLKTPTKMKQMTLLDMAKGTQKMTAPRNSG 458
Db ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
459 DKLLK-----QEMKSLAFKAKLR----- 479
QY 459 GTPRTSSKPHKLPPAALHLIAYKENKDRKDSKALSVCISKARTALLSSEDRARLPEL 518
Db ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
480 -----EKADALEAKKEKEDKEKR-----BEL 502
QY 519 RSLVQRYELLEHKRWASMSSEQRKYLYKKREELKKLKAKEKEREKEMLEKOK 578
Db ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
503 KXIVBEE-----RLKKEEKERLKVEREKEREKERE--EKRYKVEYLKQWSKPR 549
QY 579 RYEDQELTG-KNLPAFLVDTPGELPNTLFGDVAWVVEFLSCYSGLLLPDAQYP--ITAV 635
Db ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
550 --EDMECDLKLPEPTPVKT--RUPPEIFGDMVLEFLNAPGELFDLQDEFFPGVTL 605
QY 636 SLMEAL-SADKGG-----FLYLNRLVLLQTLQDEIAE-----DYGELGKMLSEIPLT 684
Db ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
606 VLEBALVNGDSGGLCELLFFFLTAIFAIAES--EEVAKELQTLADTKGCSLSKSLDLD 663
QY 685 LHSVSELVRLCRRSDVQBSGSDTD-----DNKDSAAFEDNEVDQEFLEKLETSEF 737
Db ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
664 SCTLSEILRLHLASGADVTSANAKRYQKRGFGFATDDACMELRLSNPLVKKLSSTSV 723
QY 738 FELTSEKLOILTALCHRLTMYYSVD-----HMETROQMSA 774
Db ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
724 YDLTPGKVKIHLALCGKLLTLVSTNDFIEDYVDILROAKQBFRELKABQHKREKEAAA 783
QY 775 ELWKEKRLAVLEENDKKAERKQKREMEAKN-----KENGKVENGLGHTDRKKRIV 825
Db ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
784 RTRKKEEKLKEQEQKMKKEQKLEKDEQNRSTADISICEEREDPDTIESKDTQK-- 841
QY 826 KFEQVDTBAEDMISAVKSRRLAIQAK-----KERETQE-----REMKVLERQA-E 872
Db ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
842 ELDQDMFTDEDDPGSHKRGK----RGKRGONGFKEFTREQEINCVCVTRELLTADEEALK 897
QY 873 EERIRKHAAAQKAFQEGIAKAKLVNRPTPIGTDNRHNYWFLFSDEVPGLFIEK---GWV 929
```

```
Db 898 QEHQKEKELLEK-IQSAIACTNIF-----PLGRDRMYRYWIF-PSIFGLFIEEDYSGLT 951
QY 930 HDSI-----DYRFNHCKDHTVSGDEYCPRSKKAN-LGKNASMTQHGTTATEVAVETTT 983
Db ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
952 EDMLLPSPSPQNNVQSDPQVS-----TKTGEPLMSEBTSINIDQG-PRDHSVOLPK 1002
QY 984 PQQONLWFLCDSQKELDELLNCLHPQIGIRBSOLKERL--EK-----RYQDIHSH 1033
Db ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1003 PVHKPNRWCFFSSCBQDOLIEALNSGRHSALKETLLOEKSRICAQLARFSE--EKPH 1060
QY 1034 LARKENLGLK-----SCDGNQ-----ELNLFRLSDLIEVATRLQKGLGVVEET- 1077
Db ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1061 PSDKPPPSKPTYSRGRSSNAYDPSQCAEQLELRDLDFLLDIEDRIYQGTGALIKVTD 1120
QY 1078 -----SEFEA---RVISLEKL-----KDFGECVIALQASVTKKFLQGMFAPKQKR 1120
Db ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1121 RHIMRSALSGRYELLSEENKENGIIKTWNDEVEEMEIDEQTKVIVK--DRLLGIKTETP 1178
QY 1121 KLOSEDSAKTEBVDE-----EKQNVBEAKVAS-----ALEKWKTAI 1156
Db ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1179 STVNASTPOSVSSVHYLAWALPQIQGIERRFLKAPLDASDSGRSVKTVLDRWRESL 1238
QY 1157 REAOTFSRMHVLGLMLDACIKWMSAENARCKVCPKGGEDDKLILCDECNKAFHLFCLRP 1216
Db ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1239 LSSASLSQVFLHLSTLDRSVIWSKILNARCKICKGDAENMVLCDGCRGHHTYCVRP 1298
QY 1217 ALYVDPDEWQCPACQATARNRSGRNYTESASEDSEDEDEDEE-----EEEEEEEE 1272
Db ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1299 KLKTVPEGDFWCEPCRPQRCRLRSFRORPSLESDSEDEVEDSMGGEDDEVDGDEEGQSE 1358
QY 1273 EDEYEV-----AGLRLRPKTI-----RGHSVIPPAARSGRRPKCKPSTHR 1314
Db ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1359 EEEYEVEDDDSQEBEVSLSLPRGPQVRLPVKTRGLKSSFSFGQQEQEPGRYPSRSQ 1418
QY 1315 SOPK-----APPVD-----DAEYDEL----- 1330
Db ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1419 STPKTVTSKTSRSLRKINSAPPTETKSLRIASRSTRHSHGFLQADVFVVELLSPRKRG 1478
QY 1331 -----VLQTRSSRQSL----- 1343
Db ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1479 RKSANNTPENSFPNFRVIATKSEQSRVSIASKLSQESSEKRCRKRQSPSPVPT 1538
QY 1344 -----ELQKCEELHKIVKYRFSFPPEPVTREDAEYDVITHPMDFTVQ 1390
Db ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1539 LGRRSSGQGGVHELUSAPEQLVVLVRDSDSPFLKLSKIQVDPDYDIKKPIALNIIR 1598
QY 1391 NKCSGSGYSRVQEFITDMKQVFTNAEVYVNCRGS 1423
Db ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1599 EKVNKCEYKLASEFIDDIELMFSCNCFEYNPNT 1631
```

```
RESULT 13
US-10-376-537-1
; Sequence 1, Application US/10376537
; Publication No. US20030224405A1
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042001
; CURRENT APPLICATION NUMBER: US/10/376,537
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US/09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
```

; LENGTH: 1674	
; TYPE: PRT	
; ORGANISM: Homo sapiens	
US-10-376-537-1	
Query Match	
Best Local Similarity 23.1%; Score 929.5; DB 15; Length 1674;	
Matches 395; Conservative 264; Mismatches 529; Indels 525; Gaps 67;	
QY 3	PLIGRPF---PLVNPLGCEBPFPTIPHTQAFRTREYVEARLERYSERIWTCKSTGSS 58
DB 152	PLLHRKPFYQRPADLRDEEVFY-CKVTNEIFRHYDDFFERTILCNSLVWSCAVTGRP 210
QY 59	QUTHKAWEEBGEVABLLKEEPPAWYKLVLEMVH-----HNTASLKLVDTAWEIMTYK 114
DB 211	GLTYQALSESEKARQL--QSFP---EPLIIPVLYLTLTHRSRLHEICDDIFAYVKDXY 266
QY 115	AVCEECDFVGKEMKLVKVIKHPLEKVDDEATEKKSDGACDPSDDKXENSQTAQDHQ 174
DB 267	FVEETVEIRNNGARLQCTILEVLP-----PS---HQNGFANGHV 303
QY 175	KK---ETVVKDEGRESINDRARRSPRLPTSLKKGKRWAPPKFLPHKYDV---KLQ 227
DB 304	NSVDGETIISDDSETOS-----CSFQNGKKDAIDPL--FKYKVQPTYKKE 351
QY 228	NEDKIISVPADSLINTERPPNKEIVRYFIRHNALRAGTGENAPWYVEDELVKYKSLPS- 286
DB 352	HESAI---VKATQISRKHLSRDLKLFKQHC-----EPQGVK---IKASSLSY 399
QY 287	-----KFSDFLDPPYKMTLNTSTKKNVTSQDRKPSKK---SKTNSSLSPLNPKLWC 338
DB 400	KIAEQDFSYFFDDPPTFFISFANRRG-----RPPRIHISQEDN----- 440
QY 339	HVHLKSLSGSLPKVKNKSNKSPHEHLEMMKMSPNKLHNFHTPKKGPAPKPKGHS 398
DB 441	-VANKOTLA-----SYRSKATKER----- 458
QY 399	DKPLAKAGRSKILNGKSTGNSKSPKGLKTPKTKQKQMTLIDMAKGTQKMTAPRNSG 458
DB 459	DKLLK-----QEBMKSLAFKAKLKR----- 479
QY 459	GTPRTSKPHKLPPAALHLIAYYKENDREKRSALS CVISKTARLLSSEDAKLPEEL 518
DB 480	-----EKADALEAKKEDKEDKRR-----EEL 502
QY 519	RLSVOKRYELLEHKKEWASMSSEQRKEYLKKREELKKLKSKEKEREKEMLERLEKQ 578
DB 503	KKIVBEE-----RLKKKEERUKEREKEREKERE-----EKRYKVEYLKQWSKPR 549
QY 579	RYEDQELTG--KNLPAPRLVDTPEGLPNTLFGDVAMVVEFLSCYSGILLPDAQYP--ITAV 635
DB 550	--EDMECDLDELPEPTPVKT--RUPPEIFGDALMVLEFLNAPGELFDLQDEFPDGVTL 605
QY 636	SLMEAL--SADKGG-----FLYNRLVLLLOTLQDEIAE-----DYBELGKMLSEIPT 684
DB 606	VLEELVGNDSGCLLFFFLTAIFAIAE--EEEVAKEQLTDADTKGCSLKSOLD 663
QY 685	LHSVELVLCRLRRSDVQESGSDTD-----DNKDSAAPEEDNEVDQEFLEKLETSF 737
DB 664	SCTLSEILRLHLASGADVTSANAKRYOKRGGFATDDACMELRLSNPLVKLSSTSV 723
QY 738	FELTSEKQLIALTALCHRLIMTYSVQD-----HMETQOQMSA 774
DB 724	YDLTQEKWKILHALCGKLLTLVSTDFIEDYVDILROAKQEFRELKAEQHKKEHEAAA 783
QY 775	ELWKERLAVLKEENDKRAEKQKREMEAKN-----KENGKVENGKGTDRKKRIV 825
DB 784	RIRKKEELKQEQKQKQKLEDEQNRSTADISIGEEREDPDTSIESKDTQK-- 841
QY 826	KPEQVDTAEADMI SAKVSRLLATQAK-----KEREIQE-----REMKVLERQA-E 872
DB 842	ELDQDMFTDEDDPGSHKGR-----RGRKGQNGFKEFTRQEQINCVTRELLTADBEELK 897

QY 873	ERIRRHKAAAKAFQEGIAKALVNRRTPIGTRDNRHNRWLFSDVPGLFIEK---GWV 929
DB 898	QEHQKEKELLEK-IQSAIACTNIF-----PLGRDRMYRRYWF-PSIPGLFIEDYSGLT 951
QY 930	HHSI-----DYRFNHCKDHTVSGDEDYCFPRSKAN-LGKNASMTQHGCTATEVAVETTT 983
DB 952	EDMLLRPRPSFQNNVQSQDPQVS-----TKTGPELMSESTSNIDQG-PROHSVOLPK 1002
QY 984	PRQGNLWFLCDSQKDELINCLNLPQGIRESOLKERL--EK-----RYQDIHSH 1033
DB 1003	PVHKPNRWCFYSSCQDLQIEALNSRGHRESALKETLLQEKSRICAQLARFSE--EKFH 1060
QY 1034	LARKNPLGLK-----SCDGNQ-----ELLNPLRSLDIEVATRLKGGGUYVEET- 1077
DB 1061	FSDKQPDQSKPTYSRGRSSNAYDPSQCAKEQLRLDRFLDLIEDRIYQGTGALCAKVD 1120
QY 1078	-----SEEA---RVISLEKL-----KDFGECVIALQASVILKFLQGFNAPKQKR 1120
DB 1121	RHIMWSALESGRYELLSEENKENGITVNEVDVEIMEIDQTKVIVK--DRLLGIXTETP 1178
QY 1121	KLQSEDSAKTEBVE-----EKMVVEAKVAS-----ALEKWKTAI 1156
DB 1179	STVSTNASTPQSVSVVHYLAMALFQIEQIEERRFLKAPLDASDSGRSVKTVLDRWRESL 1238
QY 1157	REAQTFSRMHVLLGLMDACIKWDMSAENARCKVCPKKGEDDKLILCDECNKAPHLFCLRP 1216
DB 1239	LSSASLSQVFLHLSTLDRSVIWSKSLNARCKICKKGAENMVLCDGCDRGHHTVCVRP 1298
QY 1217	ALYEVPDGEWQCPACQATARNRSGRNYTESASEDSDEDSDEEE-----EEEEEE 1272
DB 1299	KLKTVPEGDFWFCRCPQRCRRLSFRQRPSEDESDEVEDSMGDEDDVDGDEEGQSEE 1358
QY 1273	EDEV-----AGLRRLRPKTI-----RGKHSVIPPAARSGRRPGKPHSTR 1314
DB 1359	EYEVQEDDDQEEVEVSLPKRGPOVRLPVKTRGLKSSSSSRQOQEGEYRPSRQ 1418
QY 1315	SQPK-----APPVD-----DAEDEL----- 1330
DB 1419	STPKTVSSKTSRSLKINSAPPTETKSLRIASRSTRHSHGLQADVFVVELLSPRKRKG 1478
QY 1331	-----VLQTKRSRQSL----- 1343
DB 1479	RKSANNTPENSPNFPNFRVIATKSSEQSRSVNIASKLSQESSEKRCRKQSPSPVT 1538
QY 1344	-----ELQKCEILHKIVKRYFSFPREPVRDDEAYDYDVTTHPMDFTQV 1390
DB 1539	LGRRSSRGQGVHLSAFEQLVVELVRHDDSWPFLKLVSKIQVPDIYDIILKPIALNIIR 1598
QY 1391	NKSCGYSRSVQEFITDMKQVFTNAEVVYNCGRS 1423
DB 1599	EKVNCKEYKLASEFIDDIIELMFNCFCYFNPRNT 1631

RESULT 14

US-10-702-148-1
; Sequence 1, Application US/10702148
; Publication No. US20040063145A1
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042002
; CURRENT APPLICATION NUMBER: US/10/702,148
; CURRENT FILING DATE: 2003-11-05
; PRIOR APPLICATION NUMBER: PRIOR APPLICATION NUMBER: US/09/839,479
; PRIOR FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18

; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 69
; LENGTH: 1673
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-839-479-69

Query Match 11.5%; Score 918; DB 9; Length 1673;
Best Local Similarity 22.8%; Pred. No. 4.2e-43;
Matches 391; Conservative 263; Mismatches 531; Indels 530; Gaps 64;
QY 3 PLIGRKP-----PLVNPLPGEPPFTPIPHQEAFTREEYAEKLYSERIWTCKTGSS 58
DB 152 PLLHRKPPVRQKPPADLRPDPEEVY-CKVTNEIFRHYDDPFERTILCNLSLVMSCAVTRP 210
QY 59 QLTHKAWEEQEAELKKEEPANWEKLVLEMH-----HNTASLEKLVDTAWLEIMTKY 114
DB 211 GLTYQALESEKKARQNL-QSFP---EPLIIPVLYTSLTHRSRLHEICDDIFAYVKDRY 266
QY 115 AVGEEDCFEYGEKMLKVKIVKIHLEKVDDEATEKKSDGACDPSGSDKENSQIAQDHQ 174
DB 267 FVZETVEIRNNGARLQCTILEVP-----PS-----HQNGFANGHV 303
QY 175 KK---ETVVKDEGRESINDRARRSPKLPSTLKKGERKWAAPPKFLPHKYDV---KLQ 227
DB 304 NSVDGETIISDSDSETQS-----CSFQNGKKDAIDPLL-PKYVQPTKKEL 351
QY 228 NEDKILSNYPADSLIRTERPNKEIVRYFIRHNALRAGTENAPWVVEDELVKYSLPS- 286
DB 352 HESAI---VKATOISRRKHLFSRDKLFLKQHC-----EPOEGVIK---IKASSLTY 399
QY 287 -----KFSQPLLDPKYMTINPSTKKNVTSKPKK---SKTONSSLSPLNPKLWC 338
DB 400 KIAEQDFSYFPDDPTFFISPANRRG-----RPPKRIHISQEDN----- 440
QY 339 HVHLKSLGSLPLKVNKSNKSPBEBHEBMMQWSPNKLHTNFHPKXGPPAKPKGHS 398
DB 441 -VANKQTLA-----SYRSKATKER----- 458
QY 399 DKPLKAKGSKGILNGQKSTGNSKSPKGLKTKTKWQMTLLDMAKGTQKMTAPRNSG 458
DB 459 DKLLK-----QEEKSLAFKAKLKR----- 479
QY 459 GTPRTSSKPKHLPALHLLIAYKENKOREDKRSALSVCISKTARLLSSDDRARLPEEL 518
DB 480 -----EKADALEAKKEKEDKCKR-----EEL 502
QY 519 RSLVQKRYELLEHKRWASMSSEQRKEYLKKREELKKLKEKAKERREKEMLERLEKQK 578
DB 503 KKIVEEE-----RLKKKEERLKVREKEREKLEEB-----KRYVEYKQWSK 546
QY 579 RYEDQELTG-KNLPAPRLVDTPGLNTPFLGDMVAVVFLSCVSGLLLPDAQVP---ITAV 635
DB 547 PREDMECDLKEPPEPTVKTL---RLPPEIFGDALMVLEFLNARFELDLQDEFPDGVTL 604
QY 636 SLMEAL-SADKGG-----FLYNRLVILLQTLLODEIAE---DYGELGMKLSIPLT 684
DB 605 VLEELVGNDSGLCELLFFFTAFQIAEB---EEVAKQELTDADTKGCSLSKSLDLD 662
QY 685 LHSVSELVRLCLRRSDVQSESGSDTD-----DNKSDAAAFEDNEVDQEFLEKLETSF 737
DB 663 SCTLSEILRLHILASGADVTSANAKRYOKRGGFATDADCMELRLSNPSLVKLSSTSV 722
QY 738 FELTSEKQILTALCHRLIMTYSD-----HMETROQMSA 774
DB 723 YDLTPGKMKILHALCGKLLTLVSTDFIEDYVDILRQAKQEBFRELKAEQHRKEBEAAA 782
QY 775 ELWKEBLAVLKENDKKRAEKOKRKEWAKN-----KENGKVENGLGKTDKRIKIV 825
DB 783 RIRKREELKLEQEQMKKEQKLEKEDQBNSTADISIGEEREDPDTIESKOTEQK--- 840

QY 826 KFEPOVTEADMISAVKRRRLAIAQK-----KEREIQE-----REMKVKLROA-E 872
DB 841 ELDQDMFTEDDDPGSHKGR-----RGKRGQGFKEFTREQINCVTRELLTADEEBALK 896
QY 873 EERIRKHAAAEKAFQEGIAKAKLVMRPTPIGTDNRHNRWYLFSDVDEVPGLFIEK---GWV 929
DB 897 QEHQRKEKLEK-LQSAIACTNIF-----PLGRDRMYRYWIF-PSLPLGFIEDYSGLT 950
QY 930 HHSI-----DYRFNHHCKDHTVSGDEDYCPRSKKN-LGKNASMTQHGATVAVETTT 983
DB 951 EDHLLPRPSFQNNVQSQDPQVS-----TKTGELPMSESTSNIDQG-PRHVSQVLPK 1001
QY 984 PKQGNLWFLCDSQKDELLNCLHPQGIRESOLKERL----- 1021
DB 1002 PVHKPNRWCFYSCEQLDLIEALNSRGHRESALKETLQEKSRICAQLARFSEKPHFS 1061
QY 1022 EKRYQDIHSHILARKPNLGLKSCDQNO-----ELNFLRSLDIETVATLQKGLGVVEE 1076
DB 1062 DKRQPSKPTYSGRSSN---AYDPSQMAEKQLELRDLRDLIEDRIYQGTGAIKV 1117
QY 1077 T-----SEFEA---RVISLEKL-----KDFGECVIALQASVKKFLOGFMAPKOK 1118
DB 1118 TDBHIWRSALSGRYELLSEENKENGIIKTVDVEDVEEMIDEQTKVIVK---DRLLGIKTE 1175
QY 1119 RRKLQSEDSAKTEEVDE-----EKQWVEAKVAS-----ALEKWT 1154
DB 1176 TPSTVSTNASTPQSVSSVHYLAMALFQIEQGLRERFLKAPLDASDSGRSYKTVDWRRE 1235
QY 1155 AIREACTFRSMHVLGMLDACIKWMSAENARKVCPKGEDDKLILCDECNKAPHLFCL 1214
DB 1236 SLLSSASLSQVFLHLSLTDLSRVWISKILNARCKICRKGDAENMVLCDGCDRGHTYCV 1295
QY 1215 RPALVEYVPOGEQWQCAPQATARRNSRGRNYTEESASEDEDEDEDEDEDEDEDEDEDE 1270
DB 1296 RKPLKIVPGDMFCEPCPKQRCRLSPRQPSLEDEDEDEDEDEDEDEDEDEDEDEDEGQS 1355
QY 1271 BEEDYEV-----AGLRLRPRTI-----RGKHSVIPPAARSRRPGKXPHST 1312
DB 1356 EEEVEVEQEDDSQEEBESVLPKGRGPVRLPVKTRGLKSSFSRSGQOQEPGRYPSRS 1415
QY 1313 RRSQPK-----APPVD-----DAEVDL----- 1330
DB 1416 QOSTPKTVSSKTRSLRKINSAPPTETKSLRIASRSTRHSHGHPLOADVVELLSPRKR 1475
QY 1331 -----VLQTKRSSRQSL----- 1343
DB 1476 RGRKSANNTPENSPFPNFRVIATKSSQSRSVNIATSKLSQESSESKRCKRQSPSP 1535
QY 1344 -----ELQCEBILHKIVKRYFSWPFREPVTREADEADYDVTTHPMDPQT 1388
DB 1536 VTLGRRSSRGQGVHLSAFEQLVVELVRHDDSWPFLKLVSKIQVDPDYDIKKPIALNI 1595
QY 1389 VQNKSCGYSRVSQVDFLTMKQVFTNAEVTNCRGS 1423
DB 1596 IREKVNCKEYKLASEFIDDIELMFNSNCFEYNPRT 1630
RESULT 16
US-10-376-537-70
; Sequence 70, Application US/10376537
; Publication NO. US2003022405A1
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042001
; CURRENT APPLICATION NUMBER: US/10/376,537
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US/09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24

```
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 70
; LENGTH: 1673
; TYPE: PRG
; ORGANISM: Homo sapiens
US-10-376-537-70

Query Match      11.5%; Score 918; DB 15; Length 1673;
Best Local Similarity 22.8%; Pred. No. 4.2e-43;
Matches 391; Conservative 263; Mismatches 531; Indels 530; Gaps 64;

QY 3 PLLGRKPP-----PLVNPPLGEBPPFTI PHTQAFRTREYEAERLERYSRINTCKSTGSS 58
Db 152 PLLHRRPFVRQPPADLRDDEEVFY-CVNTNIFRHYDDFFERTILCNLSVMSCAVTRGP 210
QY 59 QLTHKEAEEEOEVAELKEEPPAWYEKLVLEMVH-----HNTASLEKLVDTAWLEIMTKY 114
Db 211 GUTYQALSEKARQNL-QSPF---EPLIIPVLYLTSLTHSRHLHEICDDIFAYVKDRI 266
QY 115 AVGEEDCFVGKEMKMLXIVKIHPLEKVDDEATEKKSDGACDSPSSDKENSQIAQDHQ 174
Db 267 FVEETVEVIRNNGARLQCTILEVLP-----PS-----HQNGFANGHV 303
QY 175 KK-----ETVVKDEGRRESINDRARRSPKRLPTSLKKGERKWAPPKFLPHKYDV----KLQ 227
Db 304 NSVDGETIISDSDSETQS-----CSFQNGKKDAIDPLL-FKYKVQPTKKEL 351
QY 228 NEDKLI SNPADSLIRTPPNKEIVRYFIRNALRAGTGENAPWVEDELVKYSLPS- 286
Db 352 HESAI---VKATQISRRKHLFRDKUKFLKQHC-----EPQGVK---IKASSLSY 399
QY 287 -----KFSDFLLDPYKYMTLNPSTKTKNTGSPDRKPSK---SKTDNSSLSPNPKLWC 338
Db 400 KIAEQDFSYPDPPTFFISPANRRG-----RPPKIHISQEDN----- 440
QY 339 HVHLKKSGLSPLKVNQNSKSPBEHLEMMKMSPNKLTNFIHPKKGPPAKPGKHS 398
Db 441 -VANKOTLA-----SYRSKATKER----- 458
QY 399 DKPLAKAGRSKGLNGQKSTGNSKPKKGLKTPKTKMKQMTLLDMAKGTKYTRAPNSG 458
Db 459 DKLLK-----QEMKSLAPEKAKIKR----- 479
QY 459 GTPRTSSKPKHLLPAAHLIAIYKENKOREDKRSALS CVISKTARLLSSEDRARLPEEL 518
Db 480 -----EKADALEAKKKEKEDKKKR-----EEL 502
QY 519 RSLVQKRYELLEKKRWASMSBQRYEYLLKKREELKKLKKKAKERRERKEMLERLEKQK 578
Db 503 KKIVEBE-----RLKKKEERLKVREKEREKREBE-----RKRYVEYKQWSK 546
QY 579 RYVEDQLTG-KNLPAFLVDTPGLPNTLFGDVMVVEFLSCVSGILLPDAQVP--ITAV 635
Db 547 PREDMECDLKELPETPVKT--RLPPEIFGDALMVLEFLNAPFELDQDEFPDGVTLLE 604
QY 636 SLMEAL-SADKGG-----FLVLRVLVILLOTLQDEIAE-----DYGELGMLKLSIPT 684
Db 605 VLEELVGNDSSEGPLCELLFFFLTAIFAIAEB--EEEVAKELTDADTKGCSLKSLDLD 662
QY 685 LHSVSELVLCRLRSVQSESGSDTD-----DNKDSAAFEDNEVQDEFLEKLETSFF 737
Db 663 SCTLSEILRLHILASGADVTSANAKRYQKRGGFDTDDACMELRLSNPSLVKCLSSTSV 722
QY 738 FELTSEKILQILTALCHILMTYSVD-----HMETROOMA 774
Db 723 YDLPGERKWLHALCCKLTLVSTRDFTIEDYVDILRQAKQSFRELKACQHKREBEAAA 782
QY 775 ELWKERLAVLKEENDKRAEKQKRKEMEARN-----KENGKVENGLGKTRDKRIV 825
Db 783 RIRKREELKEQEQWKKEQKLEKEDQQRNSTADISIGEBEEDFDTSIESKDTQK-- 840
```

RESULT 17

```
US-10-702-148-69
; Sequence 69, Application US/10702148
; Publication No. US20040063145A1
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042002
; CURRENT APPLICATION NUMBER: US/10/702,148
; CURRENT FILING DATE: 2003-11-05
; PRIOR APPLICATION NUMBER: PRIOR APPLICATION NUMBER: US/09/839,479
; PRIOR FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
```

```
QY 826 KEFQVDTAEADMISAVKSRRLAIQAK-----KEREIOE-----REMKVKLERQA-E 872
Db 841 ELDQDMFTEDEDDPGSHKRGR-----RGRKGQNGFKEFTRQEQINCVCVTRRELLTADSEEAUK 896
QY 873 EERIRKHAABAKAFQEGIAKAKLVMRRTPIGTDNRNHNRYMLFSDVPGFLFIEK---GWV 929
Db 897 QEHQRKEKELEK-LQSAIACTNIP-----PLGRDRMYRYWIF-PSLPGFLFIEEDYSGLT 950
QY 930 HDISI-----DYRFNHHCKDHTVSGDEDYCPRSKKAN-LGKNASMTNQHGTATEVAVETTT 983
Db 951 EHLPLPRPSSFOFNVSQSDPQVS-----TKTGEPLMSESTSNIDQG-PRDHSVQLPK 1001
QY 984 PQGONLWFLCDSQKELDELLNCLHPQGIRESOLKEL----- 1021
Db 1002 PVHKPNRCFYSSCSQOLDIEALNRGHRRESALKETLLOEKSRICAQLARSEEEKFHS 1061
QY 1022 EKRYQDIITHSLARKPNLGLKSCDGNQ-----ELNLFRLSDLIEVATRLQKGLGYVEE 1076
Db 1062 DKRQDPSKPTYSRGRSSN-----AYDPSQWCAEKQLELRDLRDLFLDIEDRIYQGTIGAIV 1117
QY 1077 T-----SEFEA---RVISLEKL-----KQFGCVIALQASVVIKFLQGFMAKPK 1118
Db 1118 TDRHIWRSALSGRYELLSEENKENGIIKTVNEDVEEMEIDQTKVIVK--DRLLGIKTE 1175
QY 1119 RKLOSEDSAKTEEYDE-----EKKMVEEAKVAS-----ALEKWK 1154
Db 1176 TPSTVSTNASTPQSVSSVHYLAMALFOEQGLERRFLKAPLDADSGRSYKTVLDWRRE 1235
QY 1155 AIREAQTFSRMHVLMLMDACIKWDMASENARCKVCPKKGEDDKLILCDENKAFHFLCL 1214
Db 1236 SLLSASLSQVFLHLSTLDRSVIWSKSLNARCKIKRKGDAENMVLCDCDGRHHTYCV 1295
QY 1215 RPAIYEVDPGEWQCPACOPATARRNSGRNYTEESASEDEDEDEDEE-----EEEE 1270
Db 1296 RPKLIVPEGDWFCPECRPKQCRRLSPRQSPLESDEDEVEDSMGDEDEVDGDEBEGS 1355
QY 1271 EEDYEVEV-----AGLRPRKTI-----RGHGSVIPPAARSGRPPKPKHST 1312
Db 1356 EEEYEVEDEDEDSQEEBEVSLPKRGPQVRLPVTGRKLSFSFSRGGQQPGRYPSRS 1415
QY 1313 RRSQPK-----APPVD-----DAEVDL----- 1330
Db 1416 QOSTTKTVSSKTGRSLRKINSAPTEYTKSLIASRSTRHSHGHLQADVFVELLSPRKR 1475
QY 1331 -----VLQTKRSRRQSL----- 1343
Db 1476 RGRKSANNTPENSPNFPNFRVIATKSSSEQSRSVNIASKLSQESKRRCKRQSPSP 1535
QY 1344 -----ELQKCEILHKIVKYSFWPFPREPVTREADEYDVIDTHPMDFQT 1388
Db 1536 VTLGRRSRGRQGVHELSAFEOLVVELVHDDSWFLKLVSKIQVDPDYDIKKPINALNI 1595
QY 1389 VQNKSCGSYSRYSQBFLLTDMKQVFTNAEYVNCRS 1423
Db 1596 IREKYNKCEYKLASEFIDIELMFNSCFEYFNPT 1630
```



```
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1134
; LENGTH: 141
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (105)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-864-1134

Query Match      8.9%; Score 712; DB 9; Length 141;
Best Local Similarity 99.3%; Pred. No. 8.7e-33;
Matches 140; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 464 SSKPHKHLPPAALHLIAYYKKNKREDKRSALSCVISTARTLLSSEDRARLPPEELRSLVQ 523
Db 1 SSKPHKHLPPAALHLIAYYKKNKREDKRSALSCVISTARTLLSSEDRARLPPEELRSLVQ 60

QY 524 KYELLEHKRWASHEEQKEYLKKREELKKLKEKAKERREKEMLERLEKQRYEDO 583
Db 61 KYELLEHKRWASHEEQKEYLKKREELKKLKEKAKERREKEMLERLEKQRYEDO 120

QY 584 ELTGKNLPAFRVLVDTPPEGLPN 604
Db 121 ELTGKNLPAFRVLVDTPPEGLPN 141

RESULT 19
US-10-087-192-663
; Sequence 663, Application US/10087192
; Publication No. US20020182586A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: CANCER
; FILE REFERENCE: 529452000122
; CURRENT APPLICATION NUMBER: US/10/087,192
; CURRENT FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 2059
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 663
; LENGTH: 1586
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-087-192-663

Query Match      7.6%; Score 605; DB 13; Length 1586;
Best Local Similarity 20.8%; Pred. No. 2e-25;
Matches 315; Conservative 232; Mismatches 556; Indels 408; Gaps 59;

QY 137 IHPLEKVEEATE--KKSAGCDSPSSDKE-----NSSQIADHOKKETTVEK 182
Db 236 IHPDEAAEKELTSVVAENGLVLSLEEEQPELQKMGYNGSVSVSLHQEVSLVDP 295

QY 183 -----DEGRRESINDRRSRPKLPTSLKKGKRWAPKFLPHKY--DVKLQN-EDKI 232
Db 296 PTVSCLLD--PSHLPPQLEBDTPILSSEDSLEPFDLSLAAEPVSGSLYGIDDAELMGAEK 353

QY 233 -----ISNVPADSLIRTPPNKEIVRYPIRHNALRAGTGENAPVVEDE 277
Db 354 PLEGNPVISALDCPALSNNANSLSLADDSDTSASIF-----VSPTSPPVLGES 401

QY 278 LVKYSVLSKFGDFLLDPKYMTLNPTSKTKNGSPDRPKSKSKTDSNLSLSPNPKLW 337
Db 402 VLQGSPEPSPAAAF-----QTVSPA--RKNVSSAPKARADEETTGGA- 444

QY 338 CHVHLKKSLSGSPKLVKNSKNSKSPSEHLEENKMMKSNPKLHTNFHHPK----- 386
```

```
Db 445 -----VSGSGDVLK-----RRIATPEE-----VRLPLQHGWRREVRIKKGSHRWQGETWY 489
QY 387 KGPPAKPGKHSKDPKAKAGRSKGLNQCKSTGHSKSPKGL-----KTPK-TMKQM 438
Db 490 YGPGCKRMKQF--PEVIKYLRSNVVHSVRREHFSFSPMPVGDFFERDTPTEGLQWVQL 546
QY 439 TLLDMAKGTOKMTRAPRNSGGTPTSSK-PKHKL-----PAALHLIAYYKKNKORE 489
Db 547 SAEIPIRSIQAIT-----GKRGRPRNEKAKNKEVPKVRGRGRPPPKIKMPELLNKTDRNL 602
QY 490 DKRSALSCVISTARTLLSSEDRARLPPEELRSLVQKYELLEHKKEKWSMSEBQKEYLKK 549
Db 603 PKK-----LETQEILSDEDDKAKMTKKNKKMRQK-VQGESQTPVQOARKRKQDTSK 654
QY 550 -KREELKKLKEKAKE-RRKEMLERLEKQRYEDQELTG-KNLPAFRVLVDTPEG--LPN 604
Db 655 LKQKDTKKLKRLEEQOQQAILE--EMKKPTEDMCLSDHQPLPDFTRI---PGLTLSS 709
QY 605 TLFGDVAMVVELSCYSGLLPDAQYPIYAVS-LMEALSADKGGFLYLRVLVILLOTLL 663
Db 710 RAFPCLTIVRPLHSFGKVLGFDLTQVPSLGVLOEGLLCOGDSLDKVDLLVRLKKAAL 769
QY 664 ODEIAEDYGE---LGMKLSEIPLTHSVSELVRLCLRRSDYQEESEGSTDDDNKDSAAF 719
Db 770 HDPGLPPYCQSLKILGKMSIPLTRDNVSEILRCFLMA----- 808
QY 720 EDNEVQDFLEKLETSSEFFELTSEBKLOILFALCHRI-----LMTYSVQDHMETROOMSAE 775
Db 809 --YRVPSFCDLSLRTQPPQAOPPOQKAAIHLAFVHELNSSTIIINEIDKTLSESVSCRKN 866
QY 776 LW--KERLAVLKEENDKKRAEKQKEKEMAKNKGKENGKVGKGTDRKRIVKFPQV-D 832
Db 867 KWIVEGRLLRLTALAKTGRPEVME-----GAEDGLGRR-RSRIMEETSGTEE 916
QY 833 TEAEDMISAVKSRRLAIQAKKEREIOREMKV-KLERQAEERIRKHAAEAKAFQEG 891
Db 917 EEEEEHTTAVHGR--GRKEGEIDVAASSIPELERHIEKLSKR-----QLFRKKL 965
QY 892 AKAKLVMRRTPIGTDNRNRYLWFSDEVPGLFIE--KGWVHDSIDYRNFNHC-----KDH 944
Db 966 LHSSQMLRAVSLGQDQRYRRHYWVL--PYLAGIFVEGSEGSTGPFNFSAQIPCDAAAPT 1024
QY 945 TVSGDEDYCPRSKANLCKNASMNT----- 969
Db 1025 AVSEDQ---PTPSLOLLASSKPMNTPGAANPCSPVOLSTHLPCTGTPKRLSGDSEMSOS 1081
QY 970 -----QHGTAATEVAVETTTPKQGNLWFLCDSQKELDELLNCL 1007
Db 1082 PTGLGQPKRGCRPPSPKFFKQVEQH-YLTQLTAQPIPPMCSGMMWIRDP-ETLDVLLKAL 1139
QY 1008 HPQGIRESOLKERLEKRYODIHSIHLARKNPLGLKSCDGNQOELN--FLRSLDLEVATR 1065
Db 1140 HPRGIREKALHKLSK-HKDFLQEVCL-----QPLTDPFPEPNEL-----PA 1180
QY 1066 LQKGLGYVEETSEFEARVISLEKIKDFGECVIALQASVIKFLQGFMAP----- 1115
Db 1181 LEEGVMSWPKEKTYETDVLAVLQWVEELEQRVLSLQI-----RGWTCPTPDSTREDIT 1235
QY 1116 -----KQKRKRKLOSEDSAKTEE 1132
Db 1236 YCEHLPSDPEDIPWGRGREGTVPQORNNNPLDLAVMLAVLEQNVERRYLREPLWAHE 1295
QY 1133 VDEEKKMVE-----EAKVASALEKWKTAIREAQFTSRMHVLLGLMDLCAIKWD 1179
Db 1296 VVVEKALLSTPNAGPDGTSTEISYBITPRVVRQOTLERCSRSAQVCLCMGOLERSIAWE 1355
QY 1180 MSAENARCKVCKKGGDDKLLCDSNKAFLHCLRLPALYVEYDGEWOCFACOPATARN 1239
Db 1356 KSNVKTCLVCKRGNDLFLCCDGRGCHLYCHRPMEAVPEGDWFCVCL----- 1408
QY 1240 SGRGRNYTEESASEDSEDEDEEEEEEEDYEVAGLRPRKTRIGKHVSIPI- 1296
```

Db 1409 -----SOVEEYTORPGFPKRGOK-----RKSSPLTF 1436

Qy 1297 PAARSGRRPKPHSTRRSOPKAP--PVDDAEVDLVLQTKRRSRQSLQKCEILHK 1354

Db 1437 PEGDSRRR---MLSRSDSPAVPRYPEDGLSPK---RRRHSRGGHSDLTFCFIIIME 1489

Qy 1355 IVKYRFSWPREPVTDEADYDVITHPMDFTQVONKSCGYSRVSQVDFLTDMKOVFTN 1414

Db 1490 MESHDAAMPFLEVPNPRLVSGYRRVTKNPMDFSTMRLLRGGYTSSEFAADALLVFDN 1549

Qy 1415 AEVYNCRGSHV 1425

Db 1550 CQTFNEDDSEV 1560

RESULT 20

US-10-087-192-666

; Sequence 666, Application US/10087192

; Publication No. US20020182586A1

; GENERAL INFORMATION:

; APPLICANT: Morris, David W.

; APPLICANT: Engelhard, Eric K.

; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR

; FILE REFERENCE: 529452000122

; CURRENT APPLICATION NUMBER: US/10/087,192

; CURRENT FILING DATE: 2002-03-01

; PRIOR APPLICATION NUMBER: US 09/747,377

; PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: US 09/798,586

; PRIOR FILING DATE: 2001-03-02

; NUMBER OF SEQ ID NOS: 2059

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 666

; LENGTH: 1873

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-087-192-666

Query Match 6.8%; Score 540.5; DB 13; Length 1873;

Best Local Similarity 19.3%; Pred. No. 1.1e-21;

Matches 263; Conservative 183; Mismatches 432; Indels 487; Gaps 41;

Qy 415 QKSTGNSKSPXGLKTPKTKMOMTLDDMAKGTQ---KNMTRAPNSGGTPTSSKPHKHL 471

Db 616 QAITGRGRPR---NTEKAKTKEVPKVRGGRPPKVKITEL-----LNKTDNRLPKKL 666

Qy 472, PPAALHLIAYYKENDREKRSALSVCISKAR-----LLSSEPARLPEEL 518

Db 667 EA-----QETLNEEDKAK-----IAKSKKMRQKVQGEQCQTTIQOARNKRKQET 712

Qy 519 RSLVOKRYELLEHKRWASMSREORKEYLKKREELKKLKE---KAK-----563

Db 713 KSLKQKAEKKKKEKGTQKQELKEKVRKEKKEVKMEKEVTAKPACAKADTKTLAQ 772

Qy 564 ---ERREKEMJERLEKQRYEQELTG-KNLPAFLVDTPEG--LPNTLFGDVAMVVEFL 617

Db 773 RLLEERQORQMILEMKKPTEDMCLTDHQLPDFSrv---PGTLPSGAFSDCLTIVEFL 829

Qy 618 SCYSGULLLPPAQYPTAVS-LMEALSADGGFLYLNRLVLILOTLQDEIAEDYGE---673

Db 830 HSGFGVLGPDPAKDPVSLGVQEGLLCQGDLSLGEVDLLVRLKKAALHDPGFPFSCSLK 889

Qy 674 -LGMKLSLEPLTLHVSSELVRLCLRRSDVQESSEGSTDDNKSAAAFEDNEVQDFLEKL 732

Db 890 ILGEKVSLEPLTRDNVSEILRCLFMAYG-----VEPALCDRL 926

Qy 733 ETSEFFELTSBEKLIQTLTALCHRI-----LMTYSVDHMETROQMGAELW--KERLAVLKE 786

Db 927 RTQPQAOQPPQKAAVLAVLVELHNGSTLIINEIDKTLSSMSYRKNKIVGSLRLRLK- 985

Qy 787 ENDKRAEKQKKEAKNKNGKVENGLGKTRKKRIYKFPQVDTEADMSIAVSKRR 846

Db 986 ---TVLAKTGRSEVMEGP-----BECLGRRSSRIMBEETSMEEBEESIAAIVGRR 1037

Qy 847 LLAIOAKKEREIOEREMKV-KLERQAEERIRKHAABKAKAFOEGIAKAKLVMRRTPIGT 905

Db 1038 -----GRDGEVDATASSIPELEROIEKLSK-----QLFFRKLLHSSQMLRAVSLGQ 1086

Qy 906 DRNHNYWLFPSDEVPGLFIE-----KGWVHDSIDYRF-----937

Db 1087 DRYRRRYWL-PYLAFIVVEGTEGNLVPBEVINKETDSLKVAHAASLNPALFMSKMWELAG 1145

Qy 938 -----NHHCKDHTVSGDEYCPRSK-----957

Db 1146 SNTTASSPARARGRPRTKTPGSMQPRHLKS-PVRGQDSQPOQLOLQPEALHHPAOPQPO 1204

Qy 958 -----KANLQKN-----ASMTQIGTATEVAVETTTPKQG- 987

Db 1205 LQLQLQSHKGFLEQEGSPLSQSHDLQSAPLSWLSQTSQSHSSLLSSSVLTPDSSPGK 1264

Qy 988 -----QNLWFLCDSQ-----997

Db 1265 LDPAPSQPPEPEDEAESDPQALWFNISACMPCNAAPTPPAVSEDPQTPSPQOLAS 1324

Qy 998 -----997

Db 1325 SKPMNRPSAANPCSPVQFSSTPLAGLAPKRRAGDPGEMQSPGTGLQPKRRGRPPSKFFK 1384

Qy 998 -----KELDELINCLHPOGIRRESQLKERLEKYQDI 1028

Db 1385 QMEQRYLTQLTQAQVPPCMSCGWWIIRDFEMLDAMLKALHPRGIRKALHKLHKLK-HRDF 1443

Qy 1029 IHSIHL--ARKPNLGLKSCDGNQ--LLNFLRSLIEVATRLQKGGYGVETSEFARVI 1085

Db 1444 LQEVCLRPASADPIFEPRQLPAFQEGIMSPKX-----KTYETDLAVLQWVELEORVI 1497

Qy 1086 S-----LEKLDKDFGECV-----1097

Db 1498 MSDLQIRGWTCPSPDSTREDLAYCEHLSQSDQEDITWGRGRGLAPQRTKTNPLDLAVMR 1557

Qy 1098 -IALQASVIKFLQGFMAPKQKRRKLQSDSAKTEEVDEKKWVE-----EA 1143

Db 1558 LAALQONVRRVLRPLWP-----THEVLEKALLSTPNGAPEGTTTEISY 1603

Qy 1144 KVASALEKWKTAIRAQOTFSRMHVLGLMDACIKWDMSAENARCKVCPKKGEDDKLILCD 1203

Db 1604 EITPRIVWRQTLERCRAAQVCLGLQGLSIAEKSYNKVTCLVCRKGDNDFFLLCD 1663

Qy 1204 ECKNAFLHFLCLPALIYEPDGEWQCPAQOPATARRNSRGRNYTEESASEDSEDEDEE 1263

Db 1664 GCDRGCHYCHRPKMEAVPEGDMFCTVC-----LAQVGEFTQKPG 1705

Qy 1264 EEEEEEEEDYEV--AGLRRLPRKTRGKHSVIPPAAARSGRRPGKKPHSTRRSQPKAP 1320

Db 1706 FPKRGQKRSKGSYLNFSQEGDGRRRRVLLNGRES--PAAGPRYSEEGLSKSKRR-----1757

Qy 1321 PVDDAEVDLVLQTKRRSRQSLQKCEIILHKIVKRYFSWPFREPVTDRDEADYDV 1380

Db 1758 -----LSMRNHSDLTFCFIIIMEMESHDAAMPFLEVPNPRLVSGYRRII 1802

Qy 1381 THPMDFTQVONKSCGYSRVSQVDFLTDMKQVFTNAEVYNCRGSHV 1425

Db 1803 KNPMDFSTMRLRLRGYTSSEFAADALLVFDNCQTFNEDDSEV 1847

RESULT 21

US-09-839-479-21

; Sequence 21, Application US/09839479

; Publication No. US20020039779A1

; GENERAL INFORMATION:

; APPLICANT: Jones, Michael H.

; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR

; FILE REFERENCE: 06501-042002

; CURRENT APPLICATION NUMBER: US/09/839,479

; CURRENT FILING DATE: 2001-04-20

```
; PRIOR APPLICATION NUMBER: US 09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 21
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-839-479-21

Query Match      6.7%; Score 532; DB 9; Length 1972;
Best Local Similarity 18.1%; Pred. No. 3.5e-21;
Matches 354; Conservative 256; Mismatches 599; Indels 750; Gaps 68;

QY 8 KPFFLVNPLGPEPF-FTIPH-----TQAFRTREYEARELRYSERIWTCKST 55
Db 189 KPLSLVNAQKETYMKLVPSDVLKAGNKNITSESSLTSELRSKREQY-----KQA 241
QY 56 GSSQLTHKEAMEEEQVAELK-----EEFPAYWEKLVLEMVHN-----TA 97
Db 242 PPSQLKQESSKLVKVAALSNPKATSSSPA-HPKQTLNHNHPFLTALLGNHPNG 300
QY 98 SLEKLVDAWLEIMTKYAVGERCDPEV-----TQAFRTREYEARELRYSERIWTCKST 124
Db 301 VQSVIQEAPLATTTKTKWQKINENIAAASPTPSSPNVLSTSGRRTPGNTWPMP 360
QY 125 -----GKEML--KVKIVKI-----HPL-----EKVDEE 146
Db 361 PILHSQGEKAVSNVNPVKTQHSHPAKSLVEQFRGTDSDIPSSKSDSDSDEEED 420
QY 147 ATEKSDGACDPSDKENSSQIAODHOKKETVVKDEGRRESINDRARRSPKLTSLK 206
Db 421 EDEEDEDDESDSDSSESDSSESDTSGSEEDDDDDQDQDESDD-----T 467
QY 207 KGERKWAPPKFLPHYDVYKLVQNEKIIISNV--PADSLIRTERPPNKEIVRYFIRNALRA 264
Db 468 EGE-----KTSMKL-----NKTSSVKSPMSLCHSTSPNLHIATK----- 503
QY 265 GTGENAPWVVEDELVKKYSLSLPKSFDFLDLPYKYMTLNPSTKRVKNTGSPDRKPSKSK-T 323
Db 504 -----APGSAPAALCSESQSPA-----FLGTSSSTL-----TSSPHSGTSKRRRTV 544
QY 324 DNSSLSSPLNPKLWCHVHLKLSLGSPLKVKNSKNSKSP-----EEHLEMMKMWSPN--- 376
Db 545 DERELRIPLEYG-WORETRIRFGG---RLQGEVAYYPCGKKLRQYPEVIKYLBRNGIM 600
QY 377 -----KLHT--NFHIPKKGP-----PAKPKF 395
Db 601 DISRONFSPSAKIRVCGDFYEARDGQEMQWCLLKEEDVIPRIAMEGRGRPPNDRQRA 660
QY 396 KHSIDPLKAKGRSKGI-----LNGQKSTGNS 421
Db 661 REESRMRKGRPPNVGNVAEFLDNADAKLLRKLQAEIARQAAQIKLLRKLQKQEQARVA 720
QY 422 KSPK--GLKTPKTKWQMTLLDMAGTKMTKTRAPNSGGTPTSSKPKHKLPPAALHLI 479
Db 721 KEAKQQAQIAAEKRRKQEQIKIMKQEQEKIRIOO-----IRMEKEL--RAQQL 769
QY 480 AYYKENKREDKRSALSCVISKTARLLSSDEARLPEELSLVQKRYELLEHKRWASMS 539
Db 770 EAKKKKEE-----AANAKLLEAEKRIK-EKEMR---RQAVLLKHQER----- 809
QY 540 ESKRKEYLKKGEELKKKVKAKERREKEMLERLEKQKRYEDEL-----TGKNL----- 590
Db 810 ERRRQHMLMKAMEARKAAEERLKOEKREDEKRLNKERKLEQRRLLEMAKELKKNPD 869
QY 591 -----PAFLVDIPE-GLPNTLFGDVAMVVEFLSCYSGLLLPDAQYPIITAVSLMEALS 642
```

```
Db 870 MCLADQKPELPRIPLGLVLSGSTFSDFCLMVVQFLRNFPGKVLGFDVNDVPLNSVLQEG 929
QY 643 ADKGGFL-YLNRLVLILOTLQDE-IAEDY---GELGKMLSEIPLTHSVSELVRLCLR 697
Db 930 LNIQDMSGVEQDILLVRLLSAAVCDPGLITGYAKTALGECHELLNVGNVNRDNVSEILQIFM- 988
QY 698 RSDVQEESEGSDTDNNDKDSAAFEQNEVQDEFLKLETSFEFFELTSEEKLOILTALCHRL 757
Db 989 -----EACGQTELTESLTKTAFQHTTQAQKASVLAFINELA 1026
QY 758 MTYSVQDHMETRQOQMSAELWKERLAVLKEENDKKAQEKQKEMKEMAKKENGKVENG--- 814
Db 1027 CSKSVVSEIDKNIDYMSNLRDKVWV---EGKLRLRIIHAKKTGKTDGSGIDLGBEQ 1082
QY 815 --LG-----KTDKRRIVKPEPQVDTAEADWISAVKSRRLLAIOAKKERELOE--REMK 864
Db 1083 HPLGTPTPGRKRRRKGSDYDDDDDDSD-----QGDDEDEEDKEDOK 1129
QY 865 VKLERQAEERIRKHKAABEK-----AFQEGIAKAKLVMTPTIGTDNRNR 911
Db 1130 GKKTDCDEDEGDOAAVSELEKQIEKLSQOQYRRKLPDASHLSLVMPGPRYRR 1189
QY 912 YWLFSDVEVPLFIE---KGWVH-----DSIDYRNFHC 941
Db 1190 YWIL-PRCGIFVEGMESEGLEETAKEREKLLKKAESVQIKEMPEFSGDSLNCNTDHC 1248
QY 942 -----KDH- VSGDEYCPRSK-----KANLGNA----- 965
Db 1249 EQKEDLKEKDNLTFLQKPGSFKLSKLELVAKMPPPESEVMTPKPNAGANGCTLSYQNSG 1308
QY 966 --SNMTOHGTATEVAE---TMTPKQGNLWFL- 993
Db 1309 KHSLSGVOSTATQSNVEKADSNLNTGSSPGKFSPLPNDQLLKTLEKQRQ--WESL 1366
QY 994 -----CD----- 995
Db 1367 LPRFCDDTSLTHADMSTASLVTPOSQPPSKSPSPAPLAGSSAQNPVGLNPFALLPLQV 1426
QY 996 ----- 995
Db 1427 KGVSMMLGQFCGWPTGVVTSNIPFTLSVPSLGLSGLSEGNNGSFLTSNVASSSESVP 1486
QY 996 -----SOKELDELINCLHPOGIRS 1015
Db 1487 PONEKATSAQAAVEAVKAPVDFPSPKPIPEEMQFCGWIIIDPEDLKALLKVLHLGIRK 1546
QY 1016 QLKERLEKRYQDIHSHILARKPNLGLKSCDGNQELNPLRSLDIE---VATRLQKGGIG 1072
Db 1547 ALQKQIQK-HLDYITQACLKQK-DVAIIELNENEE--NOVTRDIVENNSVEEQAMEMDLS 1602
QY 1073 VYEETSEFEARVLSLEKLKDFGECVIALQASVKKFLOGFMAPKQ----- 1117
Db 1603 VLQQVEDLERRVAS-----ASLQ-----VKGWMCPEPASEREDLVYFEHKSF 1644
QY 1118 -----KRRKQSEDSAKTEEVDDEKKMVE--EAKVASALEKKW 1153
Db 1645 TKLCKEHGEFTGEDESSAHLERSKDNPLDIATVRLADLERNIERRIEEDTAPGLRVWR 1704
QY 1154 TAIREAQFTFSRMHVLLGLMDACIKWDMGAENARCKVCPKGBDDKLIILCECNKAFHPLFC 1213
Db 1705 RALSEARSAQVALCIIQQLQKSIAMEKSIKMYCYICRKGDNELLLCDGCDKGCHTYC 1764
QY 1214 LRPALYVPDGBWQCPACOPATA-----RRNSGRNVTESASE-----DSEDESD 1260
Db 1765 HRPKITTPDGDWFCFACIAGSGTLKIKLHVKGKKTNESKKGKVTLTGDTEDSDS- 1823
QY 1261 EEEEEEEEEEDYEVAGRLRPRKTIRGKHSVIPPAARSGRRPGKPHSTRRSOPKAP 1320
Db 1824 -----ASTSSSLKRGKNDLQKMKMEENTS 1847
QY 1321 PVDDAEVDLVLQTKSSRRRQSLQKCEBILHKIVKYRFSWPFPRVTRDEADYYDVI 1380
```


Db 1645 TVLCKEHDGEFTGEDESSAHALERKSDNPLDIAVTLADLERNIERREEDIAPIGLVWR 1704
QY 1154 TAIRBAQTFSRMHVLLGLMDACIKWMDGAENARCKVCPKKGBDDKLILCECNKAFHLFC 1213
Db 1705 RALSEARGAAQVALCITQOLQKSIWEKSIIMKYVCQICRGDNEELLALLCDGDKGCHTV 1764
QY 1214 LRPALYEVDPGEWOCFACOPATA-----RNSGRNYTESASE-----DSDDSD 1260
Db 1765 HRPKITITPDGDMFPCACIAKASGOTLKILHVKGKKTNESKKGKKVTLTGDTDEDS- 1823
QY 1261 EEEEEEEEEEDYEVAGLRRLRPRKTRGKHSVIPPAAARSRRRPGKKPHSTRRQPKAP 1320
Db 1824 -----ASTSSSLKGNKDLQKRKNEENTS 1847
QY 1321 PVDDAEVDELVLQTKRRSRROSLQKCEILHKKVYRFSWPFREPVTDRBAEDYDVI 1380
Db 1848 --INLSKQESFTSVKPKRDDSKDLCALCSMILTEMETHEDAWPFLPVNLKLVPGYKKVI 1905
QY 1381 THPMDPOTVONKSCSGSYRSVOEFLTDMKQVFTNAEVYN 1419
Db 1906 KKPMDPSTIREKLSGGQPNLETFTALDVLVDFDNCETFN 1944

RESULT 23

US-10-702-148-21
; Sequence 21, Application US/10702148
; Publication No. US20040063145A1
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042002
; CURRENT APPLICATION NUMBER: US/10/702,148
; CURRENT FILING DATE: 2003-11-05
; PRIOR APPLICATION NUMBER: PRIOR APPLICATION NUMBER: US/09/839,479
; PRIOR FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 09/418,710
; PRIOR FILING DATE: 1998-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 21
; LENGTH: 1972
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-702-148-21

Query Match 6.7%; Score 532; DB 15; Length 1972;
Best Local Similarity 18.1%; Pred. No. 3.5e-21;
Matches 354; Conservative 256; Mismatches 599; Indels 750; Gaps 68;

QY 8 KPFLVNLPLGGEPP-FTIPH-----TQAFRTREYEAERLERYSERIWTCKST 55
Db 189 KPFLSVNAKQETMYKMLIVPSPDLKAGNKNTSESSLTSELRSKQY-----KQA 241
QY 56 GSSQLTHKEAWEQEVAELLK-----BEFPAYEKLVLVWVHN-----TA 97
Db 242 FPSQLKQESSKSLKKVIAALSNPKATSSSPA-HPKQTLNHNHPFLTALLGNHQNG 300
QY 98 SLEKLVDTAWLEIMTKYAVGECDFEV-----124
Db 301 VTQSVIQEAPLATTTKWQSKINENIAAASSTPSSPNLSTSGRTPGNTQWPWPSAS 360
QY 125 -----GKEKML--KVKIVKI-----HPL-----EKVDEE 146
Db 361 PILHSQGEKAVSNVNPVKTHGHPKSLVQEOPRGTDSDIPSSKSDSDSDEDEE 420
QY 147 ATEKSDGACDSPSSDKENSSQIAQDHQKQETVVKVEDEGRRESINDRARRSPKRLPTSILK 206

Db 421 EDEDEDDDDSDSDSQSSESDNSSESDTSGSEEDDDDDKQDESDSD-----T 467
QY 207 KGERKWAAPPFLPHKYDVKLQNEDKIISNV--PADSLIRTERPPPNKEIUVYPIRHNALRA 264
Db 468 EGE-----KTSMKL--NKTTSVKSPPSMLTGHSTPRNLHIAK-----503
QY 265 GTGENAPWVVEDELVKKYSLSKSDFLDDPYKYMTLNPSTKRNKNTGSPDRKPSKSK--T 323
Db 504 -----APGSAAPALCSESQSPA-----FLGTSSSTL--TSSPHSGTSKRRRV 544
QY 324 DNSLSSPLNPKLWCHVHLKKSGLKSPKVNKSNKSP-----EHLBEMKMMSPN--- 376
Db 545 DERELRIFLEYG-WQRETRIRNFGG--RLOGEVAYIAPCGKCLQYEVIVKYLNRNGIM 600
QY 377 -----KLHT-NPHIPKKG--PAKXPG 395
Db 601 DISRDNFSFAKIRVGDYFYEARDGPQEMQWCLLKEEDVIPRAMEGRGRPPNPDORA 660
QY 396 KHSKPLKAKGRSGI-----LNGOKSTGNS 421
Db 661 REESRMRKGRPPNVGNAEFLDNADAKLLKLOAQETARQAAQIKLLRKLOKQOARVA 720
QY 422 KSPKK--GLKTPKTKQKQMTLLDMAKGTQKTRAPRNSGGTPTSTSSKPHKHLPPAALHLI 479
Db 721 KEAKQQAIAEAERKQKEQIKMKQKEIKRIQQ-----IRMEKEL--RAQOIL 769
QY 480 AYYKENKDREDKRSALSCVSIKTARLLSSEDRARLPEELRSIVQKRYELLEHKRWASMS 539
Db 770 EAKKKKKE-----AANAKLLEAKRIK-EKEMR--RQOAVLLKHQER-----809
QY 540 EEOKEYLKKREELKKLKEKAKERREKEMLEKOKRYEDQEL---TGKVL-----590
Db 810 ERRQHMMLKAMEARKKAEKBERLKQKREKLEKORRELEMAKELKKPNED 869
QY 591 -----PAFLVDTPPE-GLENTLFGDVAMVVEFLSCVSGLLLPQAQPIPTAVSLMEALS 642
Db 870 MCLADQKPLPELPRIPGLVLSGSTFSDCLMVVQVFLNFGKVLGFDVNDVNLVQLSEGL 929
QY 643 ADKGGL-YLNRVLVLIQTLLODE-IAEDY---GEIKMGLSEIPLTLHVSSELVRLCLR 697
Db 930 LNIQDSMGVQDALLVRLLSAAVCDPGLITGYKAKTALGEHLLNVGNVNDVNSEILQIFM- 988
QY 698 RSDVQEESEGSDTDNDKDSAAFEDNEVQDFLEKLETSEFFELITSEELQILTALCHRL 757
Db 989 -----EAHCGQTELTESLTKFAQHTPAQKASVLAFINELA 1026
QY 758 MTYSVQDHMETRQQMSAELWKERLAVLKEENDKKRAEKQKREKEMAKKNGKVENG--- 814
Db 1027 CSKSVSEIDKXNDYMSNLRDQWV---EGKLRLRIIHAKKTKGRDTSGLDGLBEEQ 1082
QY 815 --LG-----KTRKRRIVKPEPQVDTEADMI SAVKSRRLAIQAKKEREIQE--REMK 864
Db 1083 HPLGTPTPGRRKRRKGGSDYDDDDDDSD--QGEDEDEDEKEDQK 1129
QY 865 VKLROAEEERIRKHAEEK-----AFQGIKAKLVMRPTPIGTORNHR 911
Db 1130 GKKTIDCEDEDEGQAASVEELEKQIEKLSKQQSQYRRKCLFDASHLSRVMFGPQYRRR 1189
QY 912 YWLFSDYVPLFIE-----KGWVH-----DSIDYRFNHC 941
Db 1190 YWIL-PRCGGIFVEGMESGEGLEETAKEREKIKKAEVQIKKEEMFETSGDSLNCNTDHC 1248
QY 942 -----KDHT---VSGDEDYCPRSK-----KANLGNA-----965
Db 1249 EOKEDLKEKDNLTPLQKPGSFSKLSLLEVAKMPPESEVMTPKPNAGANGCTLQYNSG 1308
QY 966 --SMTQGTGTAWE-----TTTPKQGNLWEL- 993
Db 1309 KXSLGSVOSTATQSNVEKADSNLNTGSSGFGKPYSPLPNDQLLKTILTEKRVQ--WFSL 1366
QY 994 ----CD-----995

Db 1367 LPRTPCDDTSLTHADMS TASLVTFQSQPPSKSPPTPAPLGSSAQNVPVGLNPFALSPLOV 1426
Qy 996 -----
Db 1427 KGGVMMGLQPCGWPTGVTNIPFTLSVPSLGLSGGLSGNSFLTNSVASSKSPV 1486
Qy 996 -----
Db 1487 PONEKATSAQPAAVEAKVPDFPSPKPIPEMOFGWMRIIDPEDLKALLKVLHRLGIREK 1546
Qy 1016 QAKERLEKRYODIHSHTLARKNLGLKSCDGNQELLNFRSLDIE---VATRLQKGGIG 1072
Db 1547 ALQOQK-HLDYITQACLK-K-DVAIELNEEE---NQTRDIVENWSVEEQAMENDLS 1602
Qy 1073 YVEETSEFEARVISLEKLDGFCVIALQASVKKFLQGFMAPKQ-----1117
Db 1603 VLQOVEDLERRVAS-----ASLQ-----VKGWMCPEPASEREDLVYFEKSP 1644
Qy 1118 -----KRRKQSEDSAKTEEVDEBKQWVE---EAKVASALEKWK 1153
Db 1645 TKLCKEHGDEFTGEDESSAHALERKSDNPLDIAVTRLADLERNIERRIEDIAPGLRVMR 1704
Qy 1154 TAIREAQTSRMEVLLGMDLACIKWMSAENARCKYCPKKGEDDKLLCDECNKAFHLC 1213
Db 1705 RALSEARSAQAVALCICQLOKSAWEKSIKMKVYQICRKGDNBEILLCCDGCDCGCHTYC 1764
Qy 1214 LRPALYVDPGEWQCACQAPATA-----RRNSRGRNYTESASE-----DSEDESD 1260
Db 1765 HRPKITTPDGMFCACIAKASQTLKIKKLVHKGKTKNESKGGKVVLTGTEDBEDS- 1823
Qy 1261 EEEEEEEEEDEYEVAGRLRPRKTIRGHKSVIPPAARSGRRPGKPKHSTRRSQPKAP 1320
Db 1824 -----ASTSSSLKRGKNDLOKKEKMEENTS 1847
Qy 1321 PVDDAEVDLVLQTKSSRQSLQKCEILHKIKVYRSPWFRPVTREADEYDVI 1380
Db 1848 --INLSKQSEFTSVKPKRDDSKDLALCSMLTETHEDAWPFLPLVNLKLVPGYKKVI 1905
Qy 1381 THPMDFTQVONKSCGSYRSVQBFLTDMKOVFTNAEYVN 1419
Db 1906 KKPMDFTIREKLSGQYPNLETFALDVLRLVFNDCFTFN 1944

RESULT 24

US-09-839-479-70
; Sequence 70, Application US/09839479
; Publication No. US20020039779A1
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042002
; CURRENT APPLICATION NUMBER: US/09/839,479
; CURRENT FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 70
; LENGTH: 1876
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-839-479-70

Query Match 6.6%; Score 527.5; DB 9; Length 1876;
Best Local Similarity 19.5%; Pred. No. 5.8e-21;
Matches 268; Conservative 174; Mismatches 430; Indels 499; Gaps 44;

Qy 415 QKSTGNSKSPKGLKTPKTKMKQMTLLDMAKGTQ---KMTRAPRNSGGTPTRTSSKPKHKL 471
Db 619 QAITGKRGRPR---NTEKAKTKVEVPKVRGRGPPVKVITEL-----LNKTDNRPLKXL 669
Qy 472 PPAULHLIAYYKENDKREDKRSALSCVISKTABLLSSEDPARLPEELRSUVQKRYELLEH 531
Db 670 EA-----QSTLNEEDKAKTAKSKKQMRQKVQVQEGCLTTTIOGARNKFKOSTKSLKH 720
Qy 532 KRWASMSBEORKEYLKKREELKKLKEKAKERREKE-----M 570
Db 721 KE---AKKSKAKERKTKQEKLEKVKREKKKVKKEKEEVTYKAPACKADTKLATQRR 778
Qy 571 LERLEKO-----KRYEDQELTG-KNLPAPRLVDTPEG--LPNTLFGDVAMVVEFLSC 619
Db 779 LEERQKQOQMLEBMMKPTEDMCLTDHQLPDPFSRV---PGLTLPFGAFSDCLTIVFLHS 835
Qy 620 YSGLLLPDAQYPTAVS-LMEALSADKGGFLYLNRVILVILLQTLLODETAEDYGE---L 674
Db 836 PGKVLGDFPAKQVPSLGVLOEGLLCOGDSLGEVDLLVRLKAAHLDPGPPSYCQSKKIL 895
Qy 675 GMLKSEIPLTLHSVSELVRLCLRRSDVQEESESGDTDNDKDSAAAFEDNEVQDEFLEKLET 734
Db 896 GEKVSIPITRNVNSILRCFLMAYG-----VEPALCDRLRT 932
Qy 735 SEFFELTSEBKLIQILTALCHRI---LMTYSVQDHMETRQOMSABLW--KERLAVLKEEN 788
Db 933 QPFOAQPPQOKAAVLAPFVHELNGSTLIINEIDKTLSSSYRKNKMWIVEGLRLRK--- 989
Qy 789 DKGRABKQKQKEMAKKNGKENGKVGNGIKTDRKRIVKPEPQVDTEADMI SAVKRRLL 848
Db 990 -TVLAKRTGRSEVEM-----GRPEELGRRRSRI MEETSGMBEEEEEESIAAVPGR-- 1041
Qy 849 AIQAKKERIIQEREMKV-KLERQABEERIRKHKAAAEKAFQEGIAKAKLVMRTPITGTR 907
Db 1042 ---GRDGEVDATASSIPELEQIEKLSK- ---QLFRKKLLSHSSOMLRAVSLQODR 1092
Qy 908 NHRNYW-----LFSDEV-----PGLFTIEK----- 926
Db 1093 YRRRYWVLPYLAGI FVEGTEGNLVPEEVIKKTETDSLKVAHAASLNALFSMKMELAGSNT 1152
Qy 927 -----GWV-----HDSI----- 926
Db 1153 TASSPARASRPLKTKPGMGPREFKSPVVRGQDSEQPQALQPEAQLHVAQPQPLQLQ 1212
Qy 927 -----HDSI----- 933
Db 1213 LOSHKGFLQEGSPUSLQSGQHDLSQSAFLSWLSQTHSSLSLSSSVLTDPDSSPGKLD PAP 1272
Qy 934 -----DYRFNHHCK-----DHTVSGDEYDCPRSKKANLGKNA 965
Db 1273 SQPPEPEPEDEAESSPDLQAFWNISAOQPCNAAPTPPLAVSBDQ---PTPSPQQLASSK 1329
Qy 966 SWN-----T 969
Db 1330 PMNRPSAANPCSPVQFSTPLAGLAPKRRAGDPGEMPOSQPTGLGQPKRRGRPPSKFKQM 1389
Qy 970 QHGTA TEAVETTPKQGNLWFLCDSQKELDELNCLHPQGIRESQLEKRYKQDII 1029
Db 1390 EORVLTQTAQVPPPEMCGSWWIIDPEM-LDAMLKALHPRGIREKALHHLNKH-HRDFL 1447
Qy 1030 HSIHLARKENLGLKCDGNQELLNFRSLDIEVATRL---QKG-----GLGY 1073
Db 1448 QEVCL--RPS-----ADPIPEPQLPAFOGIMSWSPEKTYETDLAV 1488
Qy 1074 VEETSEFEARVIS-----LEKUKDFGECV----- 1097
Db 1489 LQWVEELEQORVIMSDIQRGWTCPSPDSTREDLAYCEHLSQSDEDITWRGPGREGIAPQR 1548
Qy 1098 -----IALQASVIKKFQGFMAKQ---KRRKQSEDSA---KTEEVDBEK 1137
Db 1549 KTTNPDLDAVMRLAALAEQNVKRRYUREPLWPTHEWVLEKALISTPNGAPEGITTEISYE- 1607
Qy 1138 KMVEBAKVASALEKWKTAIREAQTSRMRHVLLGMDLACIKWMSAENARCKVCPKKGEDD 1197

Db 1608 -----IPRIRIWRQTLQRCESAAHVCLCLGHLERSIAWEKSNVKNVTCLVCRGDND 1660
QY 1198 KULCDECNKAFHLCFLRPALYVPGWQCPACQOPATARRNSRGRNRYTESASBDESD 1257
Db 1661 FULLCDGCRGCHYCHRPKMAVPEGDFCTVC-----LAQOVEGE 1702
QY 1258 ESDEBEEREEEDYEV---AGLRPRKTIIRKSHSVIPPAASGRPRPKKPHSTR 1314
Db 1703 FTQKPGFKRGKRGSGYSLNFSGDRRRRVLKGRES---PAA-----GPRYSEERL 1753
QY 1315 SQKAPPVDDAEVDLVLQTKSSRRQSLQKCEILHKIVKYRFSWPFREPVTDRDEAE 1374
Db 1754 SPSKERPL-----SMNHSDLTFCFILLMEMESHDAAMPFLFVPNPLVS 1799
QY 1375 DYVDVITHPMDPQTQVONKSCGSYRSVOEFLTDMKQVFTNAEYVNCGRSHV 1425
Db 1800 GYRRIIKNPMDFTWRERLLRGYTSSEFAADALLVFDNCQTFNEDDSEV 1850

RESULT 25
US-10-376-537-71
; Sequence 71, Application US/10376537
; Publication No. US20030224405A1
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042001
; CURRENT APPLICATION NUMBER: US/10/376,537
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US/09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 71
; LENGTH: 1876
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-376-537-71

Query Match 6.6%; Score 527.5; DB 15; Length 1876;
Best Local Similarity 19.5%; Pred. No. 5.8e-21;
Matches 268; Conservative 174; Mismatches 430; Indels 499; Gaps 44;

QY 415 QKSTGNSKSPKGLKTPKTKMQLLLDMAKGTQ---KMTAPRNSGGTPTSTSSPKHKL 471
Db 619 QAITGRGRPR---NTEKATKEVPKVRGRGRPPKVKITEL-----LNKTDNRPLKKL 669
QY 472 PPAALHLIAYKENKDREKRSALSCVISTKARLLSSDRARLPPELRSLVOKRYELLEH 531
Db 670 EA-----QETLNEEDKAKIAKSKKOROKVORGECLTTIQOARNKFKOETKSLKH 720
QY 532 KKRWASMSBOREKYLKKREELKKLKEKAKERREKE-----M 570
Db 721 KE---AKKSKAEKERKTKQEKLEKVKKEKVKKEEYVTKAPACKADKTLATQRR 778
QY 571 LERLEKQ-----KRYEDQELTG--KNLPAPRLVDTPEG--LPNTLFGDVAWVFLSC 619
Db 779 LBERQKQMILEBMKXPTEDMCLTDHQLPDPFSRV---PGLTLPSGAFSDCLTIVEFLHS 835
QY 620 YSGLLPDAQYPTAVS--LWEALSADKGGFLNRLVILQLTLODETAEDYGE---L 674
Db 836 PGKVLGDFPAKGVPSLGVLOEGLLCOGDSLGEVQDILLVRLKAAHLDPGPSPYCSQSKIL 895
QY 675 GMLKSEIPLTLHSVSELVRLCLRRSDVQBESEGSDDTDNKNDSAAAFEDNEVDQEFLEKLET 734
Db 896 GEKVSIEPLTRDNVSIILRCFLMAYG-----VEPALCDRLAT 932

RESULT 26
US-10-702-148-70
; Sequence 70, Application US/10702148

QY 735 SEFFELTSEELQILITALCHRI---LMTYSVQDHMETRQQMSAELW---KERLAVLKEEN 788
Db 933 QPFQAPQPOQAAVLAFFVPHELNGSTLIINEIDKTESMSSYRKWKWIVGEGRLRK--- 989
QY 789 DKRAEKQKRKEMEAKNKENGKVENGLKTKRKRIIVKFEPOVDTEADMTISAVKSRLL 848
Db 990 -TVLAKRTGRSEVEM-----GRPECLGRSSSRIMEETSGWEEEBEESIAAVPGR-- 1041
QY 849 AIQAKKERIIOEREMKV-KLROAEEERIIRKHAEEAKAFQEGIAKAKLWMRTPIGTR 907
Db 1042 ---GRDGEVDATASSIPELERQIEKLSKR-----QLFFRKLLHSSQMLRAVSLGODR 1092
QY 908 NHRNRY-----LFSDEV-----PGLFIEK----- 926
Db 1093 YRRRYVWLPYLAGIPVEGTEGNLVEEVIKKETDLSKVAHAASLNPALFSPKDELAGSNT 1152
QY 927 ----- 926
Db 1153 TASSPARARSRLTKPGMQPREFKSPVRGODSEQOALOPEAQLHVPAPQPOLQLO 1212
QY 927 -----GWV-----HDSI----- 933
Db 1213 LQSHKGFLEQEGSLSLGQSQHDLSQSAFLSMLSTQHSLLSSSLSSVLTPTDSSPGKLDPA 1272
QY 934 -----DYRFNHCK-----DHTVSGDEDYCPRSKKANLGKNA 965
Db 1273 SOPPEEPFDEABSSPDLOAFWFNISAOQMPNAAPTPPLAVSEDO---PTSPQOLASSK 1329
QY 966 SMN----- 969
Db 1330 PMNRPSAANPCSPVQFSSTPLAGLAPKRRAGDPGEMQSPOTCLGQPKRGRPPSKFKQM 1389
QY 970 QHGTAIEVAVETTPKOGONLWFLCDSQKELDELNCLHPQGIRESQKLEKRYQDII 1029
Db 1390 EQRVLTLTAQVPPEMCSGWWIPDPWM-LDAMLKALHPRGIREKALHKLHNK-HRDPL 1447
QY 1030 HSIHLARKPNLGKSCDGNQELLNRLSDLIEVATRL---QKG-----GLGY 1073
Db 1448 QEVCL--RPS-----ADPIPEPQLPAPQOEGIMSWSPKETYETDLAV 1488
QY 1074 VEETSEFEARVIS-----LEKLKDFGECV----- 1097
Db 1489 LQWVEELQQRVIMSDLOIRGWTCPSDSTREDLAYCEHLSDSQEDITWRGPGREG LAPQR 1548
QY 1098 -----IALQASVIKFFLOGFMAPKQ-----KRRKLQSEDSA---KTEEVDEBK 1137
Db 1549 KTTNPLDLAVMRLAALAEQNVKRRYLRPLWPTHEWVLEKALLSTPNGAPEGTTTISYB- 1607
QY 1138 KMVEBAKVASALEKWKTAIREAQTFSRMHVLLGMLDACIKWDMSAENARCKVCPKKGEDD 1197
Db 1608 -----ITPRIRIWRQTLQRCESAAHVCLCLGHLERSIAWEKSNVKNVTCLVCRGDND 1660
QY 1198 KLILCDECNKAFHLCFLRPALYVPGWQCPACQOPATARRNSRGRNRYTESASBDESD 1257
Db 1661 FULLCDGCRGCHYCHRPKMAVPEGDFCTVC-----LAQOVEGE 1702
QY 1258 ESDEBEEREEEDYEV---AGLRPRKTIIRKSHSVIPPAASGRPRPKKPHSTR 1314
Db 1703 FTQKPGFKRGKRGSGYSLNFSGDRRRRVLKGRES---PAA-----GPRYSEERL 1753
QY 1315 SQKAPPVDDAEVDLVLQTKSSRRQSLQKCEILHKIVKYRFSWPFREPVTDRDEAE 1374
Db 1754 SPSKERPL-----SMNHSDLTFCFILLMEMESHDAAMPFLFVPNPLVS 1799
QY 1375 DYVDVITHPMDPQTQVONKSCGSYRSVOEFLTDMKQVFTNAEYVNCGRSHV 1425
Db 1800 GYRRIIKNPMDFTWRERLLRGYTSSEFAADALLVFDNCQTFNEDDSEV 1850

Db 721 KE--AKKSAEKEGKTKGKKEKVKRKECKKVMKEKEEVTYKAKPACKADKTLATQR 778
QY 570 MLERLEKQ-----KRYEDQELTG-KNLPAPFLVDTPBG--LPNTLFGDVAMVVEFLS 618
Db 779 RLEERQKQOMILEEMKKPTEDMCLTDHQLPDSRV---PGLTLPFGAFSDCLTIVEFLH 835
QY 619 CYSGLLLPDAQYPIITAVS-LMEALSADKGGFLYLNRLVILLQTLLODEIADYGE---- 673
Db 836 SFGKVLGDFPAKQVPSLGVLEQGLLCOGSLGEVQDILLVRLKAAALHDPGFPSPQSCLKI 895
QY 674 LGMKLSIPLTLHSVSELVRLCLRRSDVQEESESGSDTDNKKDSAAFEDNEVDQEFLEKLE 733
Db 896 LGKVSIPITRONVNSILRCFLMAYG-----VXPALCDRLR 932
QY 734 TSEFFELTSEKQLILTALCHRI-----LMTYSVDHMETRQOMSALW--KERLAVLKEE 787
Db 933 TQFPQAPPOQAVALAPFVHELNGSTLIINEIDKTLLESMSYRKNKWIVEGLRLRK-- 990
QY 788 NDKRAEKQKREMEAKNGKENGKVGDKTKRKRIVKPEPOVDTEADQMISAVKSRRL 847
Db 991 --TVLAKRTGRSEVM-----GRPECLGRRSRRIEETSGMEEBEEBESIAAIVPGR-- 1042
QY 848 LAIQAKKEREIQEREMKV-KLROAEEERIRKHAEEKAFQEGIAKAKLVMERTDIGTD 906
Db 1043 ---GRDGEVDATASSIPELEKRIKSKR-----QJFRKKLLHSSQMLRAVSLGQD 1092
QY 907 RHNRYWLFDEVPGLFIE-----KGWVHDSID----- 934
Db 1093 RYRRYVWL-PYLAGFIVTEGTLNVPPEVVKETDSLKVAANASLNPAFLSPKMKMELAGS 1151
QY 935 -----YRFNHCKDHTVSGDEDYCPRSK----- 957
Db 1152 NTTASSPARARSPLKTKPGFMQPRHFKS-PVRGQDSEQPAQLOPEAQLHVPAPQPOL 1210
QY 958 -----KANLGKN-----ASMTOHGTATEVAVETTPKQG-- 987
Db 1211 QLQLQSHKGLEQSGPLSGSQHDLQSQSAFLSWLSQTSQSHSLSSSVLTDPDPSGKL 1270
QY 988 -----QNLWFLCDSQ----- 997
Db 1271 DPAPSQPPEPEDEAESPDLOAFWNISQAQPCNAAPTTPPLAVSEDPQTPSPQOLASS 1330
QY 998 ----- 997
Db 1331 KPMRPSAANPCSPVQVSTPLAGAPKRRAGDPGEMPOSPTGLGQPKRRGRPPSKFFKQ 1390
QY 998 -----KELDELLNCLHPQIGRESOLKLEREKRYQDII 1029
Db 1391 MEQRYLTQTAQVPPPEMCGMWWIPDPEMLDAMLKALHPRGIREKALHKLNK-HRDFL 1449
QY 1030 HSIHL--ARKPNLGLKSCDGNQ--LLNFLRSLDIEVATRLQKGLGVVETSEFEARVIS 1086
Db 1450 QEVCLPSADPIPEPQLPAFQBGIMSWPKE-----KTYETDVLAVQWELEORVIM 1503
QY 1087 -----LEKLDFGECV----- 1097
Db 1504 SDLQIRGWTCPSPDSTREDLAYCHELSQEDITWRGPGREGIAPQKTTNPLDLAVMRL 1563
QY 1098 IALQASVKKFLOGFMAPKQKRLKQSEDSAKTEEVDEBKQV-----EAK 1144
Db 1564 AALEQNVKRYLREPLWP-----THEVVELEKALLSTPNGAPEGTTTISYE 1609
QY 1145 VASALEKWKTAIREAOTFSRMHVLLGMLDACIKWDMSAENARCKVCPKKGEDDKLLCDE 1204
Db 1610 ITPRIIWIQTQRCSSAAHVCLGLHLERSIAWEKSVNKTCLVCRKGNDEFFLLCQD 1669
QY 1205 CNKAFHFLCRLPALVEVPDGEWQCPACQAPATARRNSGRNYITEESASESDESEDEE 1264
Db 1670 CORGCHYCHRPQWEAVPEGDFCTVC-----LAQQVEGEFTQKPGF 1711
QY 1265 EEEEEEEEDYEV---AGLRLRPKRTIRGKHSVIPPAARSGRRPGKKPHISTRSQKAPP 1321

Db 1712 PKGQKRKSGYSLNPFSEGDGRRRRVLLKGRES---PAA--GPR-----YSEERLSPSK-- 1759
QY 1322 VDDAEVDELVLQTKRRSRRSQSLQKCEIEILHIVKRYFSWPPRPVTRDEADYDVI 1381
Db 1760 -----RRRLSMRNHSDLTTCIEILMEMESHDAAPFXEPVNPPLVSGYRIIK 1808
QY 1382 HPMDFOVTQNKSCGSGYRSVQVEFLTMKQVFTNAEVYNCRGSHV 1425
Db 1809 NPMDFSTMRERLLRGYTSSEFPAADALLVDFDNCQTFNEDDSV 1852
RESULT 29
US-10-702-148-13
; Sequence 13, Application US/10702148
; Publication No. US20040063145A1
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042002
; CURRENT APPLICATION NUMBER: US/10/702,148
; CURRENT FILING DATE: 2003-11-05
; PRIOR APPLICATION NUMBER: PRIOR APPLICATION NUMBER: US/09/839,479
; PRIOR FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 1878
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(1878)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-10-702-148-13
Query Match 6.5%; Score 521.5; DB 15; Length 1878;
Best Local Similarity 19.4%; Pred. No. 1.3e-20;
Matches 265; Conservative 179; Mismatches 437; Indels 483; Gaps 43;
QY 415 QKSTGNSKSPKGLKTPKTKMOMTLDDMAKGTQ---KMTAPRNSGGTPTRTSSKPKHL 471
Db 619 QAITGKGRPR---NTEAKTKEVPKVRGRGRPPKVKITEL-----LNKTDNRPLKKL 669
QY 472 PPAALHLIAYYKENKDRDKRSALSCVISKTARLLSSDRARLPEELRSIVQKRYELLEH 531
Db 670 EA-----OETNEEDKAKIAKSKKQKQKQVORGECLTTIQGARNRKQKETSLLKH 720
QY 532 KIRWASMSBEQKEYLKKKKREELKKLKEKAKER---REKE----- 569
Db 721 KE--AKKSAEKEGKTKGKKEKVKRKECKKVMKEKEEVTYKAKPACKADKTLATQR 778
QY 570 MLERLEKQ-----KRYEDQELTG-KNLPAPFLVDTPBG--LPNTLFGDVAMVVEFLS 618
Db 779 RLEERQKQOMILEEMKKPTEDMCLTDHQLPDSRV---PGLTLPFGAFSDCLTIVEFLH 835
QY 619 CYSGLLLPDAQYPIITAVS-LMEALSADKGGFLYLNRLVILLQTLLODEIADYGE---- 673
Db 836 SFGKVLGDFPAKQVPSLGVLEQGLLCOGSLGEVQDILLVRLKAAALHDPGFPSPQSCLKI 895
QY 674 LGMKLSIPLTLHSVSELVRLCLRRSDVQEESESGSDTDNKKDSAAFEDNEVDQEFLEKLE 733
Db 896 LGKVSIPITRONVNSILRCFLMAYG-----VXPALCDRLR 932
QY 734 TSEFFELTSEKQLILTALCHRI-----LMTYSVDHMETRQOMSALW--KERLAVLKEE 787

```
Db 933 TOPQAOQPQQAVALPVPVHELNGSTLIINEIDKTLESMSYRNKNWIVGRLRLK-- 990
QY 788 NDKKAQKQKQKEMAKONKENGKVENGLGKTRKRIIVKPEQVQVTEADMTSAVKSRL 847
Db 991 --TVLAKRTGRSEVM-----GRPECLGRRRSSRIMBETSGWEEBEEESIAAVPGR-- 1042
QY 848 LAIOAKKEREIOEREMKV-KLEROAEEERIRKHAAAKAFOEGIAKAKLVMRRTPIGTD 906
Db 1043 -----GRDGEVDATASSPELEKQLEKSKR-----QLFFRKLLHSSQMLRAVSLGOD 1092
QY 907 RNHNRYWLFSDVPGLFIE-----KGWVHDSID----- 934
Db 1093 YVRRYVWL-PYLAGIFVEGTGNLVPBEVVKETDSLKVAHAASINPALFSKMKELAGS 1151
QY 935 -----YRNHHCKDHTVSGDEDYCPRSK----- 957
Db 1152 NTTASSPARARSLTKPGFMQPRHFKS-PVRGQDSEQPAQLQPEAQLHVPAQPOL 1210
QY 958 -----KANLGN-----ASMTQHGTEVAVETTPKQG-- 987
Db 1211 QLQLOSHKGFLEQEGSPUSLGQSHDLSQSAFLSWLSQTSQSHSLSSSVLTPDSSPGKL 1270
QY 988 -----QNLWFLCDSQ----- 997
Db 1271 DPAPSQPEEPDEAESPDLOAFWNISAQMPCNAAPTPLAVSEDOPTSPQOLASS 1330
QY 998 ----- 997
Db 1331 KPMRPSAANPCSPVQFSTPLAGLAPKRRAGDGPMPQSPGTGLQPKRRGRPPSKFFKQ 1390
QY 998 -----KELDELLNCLHPQIGRESOLKERLEKYODII 1029
Db 1391 MEQRYLTQITQAPVPEMCSGWWIIPDPEMLDAMLKALHPRGIREKALHKLK-LHRDFL 1449
QY 1030 HSIHL--ARKPNLGLKSCDGOE-LINFLRSLDIEVATRLQKGLGYVEETSEFEARVIS 1086
Db 1450 QEVCLRPSADPIFEQRLPAFOEGIMNSPKB-----KTYETDLAVLQWVELSQRVIM 1503
QY 1087 -----LEKLKDPGECV----- 1097
Db 1504 SDLQIRGWTCPSPDSTREDLAYCEHLSQSDQDITWRGPGREGLAPQRKTTNPLDLAVMRL 1563
QY 1098 IALQASVTKFLOGFMAWKQKRRKLQSDSASAKTEEVDEEKNVE-----EAK 1144
Db 1564 AALEQNVKRYUREPLMP-----THEVLEKALLTPNGAPEGTTTEISYE 1609
QY 1145 VASALEKWKTAIREAQTFSRMHVLLGMLDACIKWMSAENARCKVCPKKGDDKLILCDE 1204
Db 1610 ITPRIRINRQTLQRCRSAAHVCLGLHLERSIAWEKSVNKVTCVCRKGDNDPEFLLCDG 1669
QY 1205 CNKAFLFCRLRALYVDPGEVQCPAQCPATARRNSRGRNYTEESASEDSEDESEBEE 1264
Db 1670 CDRGCHYCHRPKMEAVPBGDFWCTVC-----LAQOVEGEFTQKPGF 1711
QY 1265 EEEEEEEEDYEV--AGLRLPRKTIRGKHSVIPPAARSGRRPGKKHSTERSOPKAPP 1321
Db 1712 PKRGQRKSGYSLNPSSEGRRRRLVLLKGRES--PAA--GPR-----YSEERLSPSK-- 1759
QY 1322 VDDAEVDELVLQTKSSRRSQSLELOKCEIILHKIVKYPSPWFRPVTTRDEAEDYDVIT 1381
Db 1760 -----RRRLSMRNHSDLTFCBIIILMEMESHDAWPFXPVNPRLVGYRRIK 1808
QY 1382 HPMDFQTVQKSCGSYRSVQBFELTDMKQVFTNAEYVNCRGSHV 1425
Db 1809 NPMDFSTMRRLRGYTTSSBFAADALLVFONCQTFENEDDSEV 1852
```

RESULT 30

```
US-09-839-479-71
; Sequence 71, Application US/09839479
; Publication No. US2002039779A1
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
```

```
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042002
; CURRENT APPLICATION NUMBER: US/09/839,479
; CURRENT FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 09/418,710
; FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 71
; LENGTH: 1969
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-839-479-71

Query Match 6.4%; Score 508.5; DB 9; Length 1969;
Best Local Similarity 18.0%; Pred. NO. 7.4e-20;
Matches 353; Conservative 246; Mismatches 601; Indels 763; Gaps 68;

QY 8 KPFLVNVLPQGEPP-FTIPH-----TQEAFTREYEYEARLERYSERIWTCKST 55
Db 189 KPLSLVNVQAKETYNKLVPSPDVLKAGNKTSESSLTSELRSKREY-----KQA 241
QY 56 GSSQLTHKEAWEEEOEVAELLK-----EEFPAYEKLVLVEMVHN-----TA 97
Db 242 FPSQLKQKQSSKSLKVVIAALSNPKATSSSPA-HPKQTLNHNPNFLTALLGNHPNG 300
QY 98 SLEKLVDTAWLEIMTKYAVGECDEEV----- 124
Db 301 VIQSVIQEAPLALTATKTKWQSKINENIAAASSTPSPVNLSTSGRTPGNOTPVMPAS 360
QY 125 -----GKEKML--KVKTIVKI-----HPL-----EKVDE 145
Db 361 PILHSQGEKAVANNVNPVKQHHSHPAKSLVEQFRGTDSDIPSSKDSNDEDEED 420
QY 146 EATEKKSQDAGCDSPSXDSENSQIAQHQKKEVVKVEDEGRRESINDRARRSPRLPTSL 205
Db 421 EDEDEDEDESDSQSSEDSNSE--SDTEGSEEDDDKQDQSDS----- 465
QY 206 KGERKWAAPPFLPHKYDVKLQNEBKIIISNVPADSLIETERPKNKEIVRYFIRHNALRAG 265
Db 466 TEGE-----KTSKMLNKTTTS--SKSPMSLTGHSTPRNLHIK----- 501
QY 266 TGENAPWVVEDELVKYSLPSKFSDFLLDPYKMTLPNSTKRNKNTGSPDRKPSKSK--TD 324
Db 502 ----APGSAPALCSESQSPA-----FIGHTSSSTL---TSSPHSGTSKRRVTD 543
QY 325 NSSLSPLNPKLVCHVHLKSLSGSPLKVKYNSKNSKSP-----EEHLEEMQMOMSPN---- 376
Db 544 ERELRLPLEYG-WQRETRIRNFGG--RLQGEVAVYAPCGKKLQYPEVIKYLRSNGIMD 599
QY 377 -----KLHT-NFHIPKPGP-----PAKKPKK 396
Db 600 ISRDNFSFSAKIRVGDYFYEARDGPOEMQWCLLKEEDVIPRIAMEGRGRPPNPDRQAR 659
QY 397 HSDKPLKAKGRSKGI-----LNQKSTGNSKS 423
Db 660 EESRMRRRKGRPPNVGNAEFLDNADAKLLRKLQAEARQAQIKLLRKLQKEQEARVAKE 719
QY 424 PKK--GLKTPKTKMKQMTLLDMAGTKQKMTAPRNSGGTTPRTSSKPHKLPPAALHLIAY 481
Db 720 AKKQQAIIWAABEKKQKEQIKHKQKEKIKRIQQ-----IRMEKEL--RAQOILEA 768
QY 482 YKENKDRDKRSALSCVLSKTAIRLLSSSEDRARLPBELSLVQKRYELLEKKRWASMSSE 541
Db 769 KKKKKEE-----AANAKLLEAEKRIK-----EREMRQQAVALKQREB-----BR 808
QY 542 QRKEYLKKKREELKKLKEKAKERKEKEMLERLEKQKRYEDQEL---TGKNL----- 590
```


Db 809 RROHMLKAMEARKAAEERLUKQKROKDEKRLNKERKUEQORLELEMAKELKKPNEDMC 868
QY 591 -----PAFRLVDTPE--GLPNTLFGDVAMVVEFLSCYSGLLLPDAQYPTAVS--LMEALSA 643
Db 869 LADOKPLPELPRIPLGLVLSGSTSDCLMVQVFLRNFQKVLGFDVNDVPSLVLQSGILL 928
QY 644 DKGGFL--YLNRLVILLQTLLOBE--IABDY---GELGMLSEIPLTLHVSSELVRLCLRR 698
Db 929 NIGDSMGVQDILLVRLLSAAVCDPGLITGYKAKTALGEHLLNVGNVNRDNVSEILQIFM-- 986
QY 699 SDVQBESESDTDNDKDSAAFEENEVDQEFLEKLETSEPFELTSEKLIQILTLCHRILM 758
Db 987 -----EHCQGTLETSLKTKAFOAHPAQK-AVLAFLINELAC 1024
QY 759 TVSQVDMETROQMSAELWKLAVLKEENDKRAEKOKKEMAKENKENGKVGGLG-- 816
Db 1025 SKSVSEIDKNIDYMSNL-----RRDKNVVEGLKRLRIIHAKTGKRDTSGGID 1074
QY 817 -----KTRDKRIRVKEFPQVDTEABDMISAVKSRRLLAIAQAKEREIOE-- 860
Db 1075 LGEEQHPLGTPTGCRKRRKGGSDYDDDDDDSD-----QGDDEDEED 1121
QY 861 -REMKVLERAEEERIRKHAEEK-----AFQGIKAKLVMRRTPIGTD 906
Db 1122 KEDQKGKTDICEDDEGQAASVELEKQIEKLSKQSQYRRKLPDASHLSRVNMGPD 1181
QY 907 RHNRYWLSFDEVPGLFIE-----KGMVH-----DSIDYR 936
Db 1182 RYRRYIWL--PRCGGLFVGESEGELEIAKEREKLLKAESVOIKEEMFETSGDSLNC 1240
QY 937 FNHHC-----KDHT---VSGDEDYCPRSK-----KANLGKNA----- 965
Db 1241 NTDHCQKEDLKEKDNINLFLQKPGFSKLSKLEVAKMPESEVMTPKPNAGANGCTLS 1300
QY 966 -----SMTQGTATEVAE-----TTTTKQGN 989
Db 1301 YQNSGHSLSGVOSTATQSNVEKADSNLNTGSSGPGFKFYSLPNDQLLTKLITERNQ- 1359
QY 990 LWFL-----CD----- 995
Db 1360 -WFSLLPRPCDDTSLTHADMSTASLVTPQSPKSPSPPTAPLGSQAQNPVGLNPPAL 1418
QY 996 ----- 995
Db 1419 SPLQVKGVSMMGLQFCGMPTGVVTSNIPFLSVPSLSGSLGSEGNCSFLTSNVASK 1478
QY 996 -----SOKELDELLNCLHPQ 1010
Db 1479 SESVPQNEKATSAQPAAVEAKVPDFPSPKPIPEEMQFGWRIIDPEDLKALLKVLHLR 1538
QY 1011 GIRESQLKERLEKRYODIHSIHLAKPNLGLKSCDGNQELLNFLRSLDIE---VATRLQ 1067
Db 1539 GIREKALQKQIQK-HUDYITQACLKNK-DVAIIELNENEE--NQVTRDIVENWSVEQAM 1594
QY 1068 KGLGVEETSEPEARVISEKLDKDFECVIALQASVIKKFLQGFMAKQ----- 1117
Db 1595 EMDLSVLQVEDLERVAS-----ASLQ-----VKGMWCPFASEREDLVFP 1636
QY 1118 -----KORKLQSEDSAKTEEVEDEKGMVE--EAKVASA 1148
Db 1637 EHKSFTKLCKEHDEFTGEDESSAHALERKSNPLDIAVTRLADLERNIERRIEEDIAPO 1696
QY 1149 LEKKWTAIREAQFTSRMVLGLMDACIKWDSNAENARKVCPKCGEDDKLILCDECNKA 1208
Db 1697 LRWRRALSEARAAQVLCIOLOQKSIWEKSIKMYCOICRKGNDRELLLCDCGDKG 1756
QY 1209 FHLFLRLPALYEPDCEWQCPACQAPATA-----RRNGRGRNYTESASE-----DSE 1255
Db 1757 CHTYCHRPKITIPDGDWFCPACIAKASGOTLKIKLHVKKKTNSKKGKVKTLTGDTY 1816
QY 1256 DDESDEEBEEEEDEYEVAGRLRPRKTIIRGKHSHVIPPAAARGRRPGKKPHSTRS 1315
|::|

Db 1817 DEBS-----ASTSSSLKRGNKDLQKRKM 1839
QY 1316 QKAPPVDDAEVDELVLQTKRRSRROSLQKCEEILHKIVKYSFWPFPREPVTREAEAD 1375
Db 1840 BENTS--INLSKQESTSVKPKKDDSKDALCSMLTETWETHEDAWPFLPNVNLKLVPG 1897
QY 1376 YVDVITHPMDFQTVQNKSCGYSYRSVOEFLTDMKQVFTNAEYV 1418
Db 1898 YKKVIKKPMDFSTIREKLSGQVNPNETFALDVLVDFDNCETF 1940

RESULT 31
US-10-376-537-72
; Sequence 72, Application US/10376537
; Publication No. US20030224405A1
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042001
; CURRENT APPLICATION NUMBER: US/10/376,537
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US/09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 72
; LENGTH: 1969
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-376-537-72

Query Match 6.4%; Score 508.5; DB 15; Length 1969;
Best Local Similarity 18.0%; Pred. No. 7.4e-20;
Matches 353; Conservative 246; Mismatches 601; Indels 763; Gaps 68;

QY 8 KPFPVLNPLUGSEPP-FTIPH-----TQEAFTREYEARELRYSERIWTCKST 55
Db 189 KPLSLVNAQKETYMKLIIVPSDVLKAGNKNTSESSLLTSELRSKREQ-----KQA 241
QY 56 GCSOLTHKEAEEQEVAELK-----EEFPAYEKLVLVEMVHN-----TA 97
Db 242 FPSQLKKQESSKSLKVIAAALSNPKATSSSPA-HPKQTLNHNHPFLTNALLGNHOPNG 300
QY 98 SLEKLVDTAWLIMTKYAVGECDFEV----- 124
Db 301 VIQSVIQEAPLALTTKTKQSKINENIAAASSTPSSPVNLSTSGRRTFQNTQPVMPAS 360
QY 125 -----GKEKML--KYKIVKI--HPL-----EKVDE 145
Db 361 PILHSQKGEKAVNNVNPVKTHHSHPAKSLVEQFRGTDIPSSKSDSEDEDEDE 420
QY 146 EATEKSDGACSPSDKENSQIAQDHQKKEVVKEDGREGRESINDRARRSRKLPTSL 205
Db 421 EDEDEDEDESDDSESSENSSE--SDTEGSEEDDDDDKQDESDS----- 465
QY 206 KKGKRWAPPKPLPHKYDVKLQNEDEKIIISNVPADSLIRTERPPNKIIVRYFIRHNALRAG 265
Db 466 TGGE-----KTSMLKNKITS--SKSPMSLTGHTSPNLHIK----- 501
QY 266 TGENAPVWVEDELKYSKSLPSKFSDFLLDPYKMTLNPTKRNKNTGSPDKPSKSK-TD 324
Db 502 ----APGSAPAALCSESQSPA-----FLGTSSTL---TSSPHSGTSKRRVTD 543
QY 325 NSSLSPLNPKLWCHVHLKKSLSGSPKLVKNKSNKSP-----EEHLEEMKMSPN---- 376
Db 544 ERELRLPLEYG-WORETRIRNFGG---RLQGEVAYAPCGKKLROYPEVYKILSRNGIMD 599
|::|


```

; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anncmax Sequence Listing Engine vers. 1.1
; SEQ ID NO 34544
; LENGTH: 69
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC005089.2
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.9
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.8
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2.5
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.4
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.72
; OTHER INFORMATION: EST HUMAN HIT: A1498634.1, EVALUATE 5.00e-32
; OTHER INFORMATION: SWISSPROT HIT: P13816, EVALUATE 1.70e-01
US-09-864-761-34544

Query Match 4.2%; Score 338; DB 9; Length 69;
Best Local Similarity 98.5%; Pred. NO.5e-12;
Matches 67; Conservative 1; Mismatches 0; Indels

Qy 124 VGKERMVKVIXHPLEKVDDEATKKSQDGCDCSPSSDKKNSSQIAQDHO
Db 1 VGKERMVKVIXHPLEKVDDEATKKSQDGCDCSPSSDKKNSSQIAQDHO
Qy 184 EGRRESIN 191
Db 61 EGRRESIS 68

RESULT 35
US-10-369-493-1586

```

Sequence 1586, Application US/10369493
Publication No. US20030233675A1
GENERAL INFORMATION:
APPLICANT: Cao, Yongwei
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xianfeng
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
FILE REFERENCE: 38-10(52052)B
CURRENT APPLICATION NUMBER: US/10/369,493
CURRENT FILING DATE: 2003-02-28
PRIOR APPLICATION NUMBER: US 60/360,039
PRIOR FILING DATE: 2002-02-21
NUMBER OF SEQ ID NOS: 47374
SEQ ID NO 1586
LENGTH: 1790
TYPE: PRT
ORGANISM: Saccharomyces cerevisiae
US-10-369-493-1586

Query Match 4.1%; Score 324; DB 15; Length 1790;
Best Local Similarity 20.2%; Pred. No. 1.8e-09;
Matches 274; Conservative 222; Mismatches 471; Indels 388; Gaps 58;

QY 17 PGSEPFPTPHPTQAEAFRTREEVEARLERYSERIWTCKSTGSSOLTHKEAWE-----EE 69
DB 719 PDEEPINKI-----SFEVEKLRQCTKLGEITSLOT--ETESTHENLTEKLIALTNEH 771
QY 70 QEVAE-----LLKSEFPAYEKVLVEMVHNHTA-SLEKLV-----DT 105
DB 772 KELDEKYQILNSHSHLKENFS-----ILETELKNVRSDLSDBMTQLRDVLETKKENQT 825
QY 106 AMLEIMTKYAVGEC--DEVEGKEMKVKIVKIHPLEKV-----DEATEKKSOG 154
DB 826 ALLEYKSTIHKBDSIKTEKGLLETILSOCKKAEDGINKVKGDLFALSREMOQAVEEN--- 882
QY 155 ACDSPSDKENSQIAQHQKTEVVKEDGREGRESINDRARRSPRLPTSLKKGKRWAP 214
DB 883 -CKNLQKEDKSN---VNHQKETSLKEDIAKIT-----913
QY 215 PKPLPHYDVKLQNEDKIISNPVADSLIRTPPNKEIVRYFIRHNALRAGTGENAPWV 274
DB 914 -----ETKAINENLEEMKIOCNLSKEKEHISKELVEYKSRFQS-----952
QY 275 EDELVKYSLPSKFSDFLLDPYKWTILNPSTRKQNTGSPDRPKSKSKYTDNSSLPLNP 334
DB 953 HDNLVAK--LTKELKS-LANNYKDMQAEKESLIKAV-----EESKNESIIQLSNLQN 1001
QY 335 KLWCHVHLKKSLSGSPKLVKNSKNK-----SPEEHLEEMMMQWSPNKLHTNPHI PKGPP 390
DB 1002 KI-----DSMS-----QKENFOIERSIEKNIEQKKTISDLE-QTKSEIISKSDS 1047
QY 391 AKKPGHSDPKLKAQRSKGIILNGQKSTGNKSPKGLKTPPKTKMOMTLDMMAKGTQKW 450
DB 1048 SKDEYESQISLLKEKET-----ATTANDENVNKI SELTKTREETEELEAAYKNL--- 1097
QY 451 TRAPRNSGGTP-RTSSKPHKHLPPAALHLIAYKENKDRKKSALSCVISTARLLSS- 508
DB 1098 ----KNELETKLTSTSKALKEVENEHL-----KEEIKOLEKEA-----TETQQLNLSL 1143
QY 509 ----EDRARLPEELRSILVOKRYELLEHKKRWAS-----MSEQRKEYLKKREEL 554
DB 1144 RANLESLEKEHEDLAALKKYEQIANKEQYNEEISQNLDEITSTQQNESIKKQNDLSL 1203
QY 555 KKKLKEKAKERREKEMLERLEKORYEDQELTGKLPAPFLVDTPBGLPNTLFGDVAMV 614
DB 1204 -----EGEVKAMKSTSEQSNLKKSEIDALNLQIKELKKNETNEASLESIKS-V 1253
QY 615 EFLSCYSGLLLPDPAQYPTAVSLME---ALSADKGGFLYLNVRVLVILLQTLLODETAEDY 671
DB 1254 ESETVVIKELQDCNFCNFEVSELDKJASEDKNS-KYLE-----LOKESEKIK 1302

QY 672 GELGMKLSPIPLTHSVSELVRLCRRSDVQSESESGSDTDDNKOSAAFEDNEVDSEFLEK 731
DB 1303 EELDAKTTELEKIQLEKITNL-----SKAKESESELRLKKTSS--BERKNAEBOLEK 1353
QY 732 LETS-EFFELTSEELKQILTALCHRLIMTYSDVQHEMETRQOQMSAELWKERLAVLKEENDK 790
DB 1354 LXNEIQIKNAQPEKERKLLNEGSTTITQEYS--EKINTLE-----DELIRLQWENEL 1403
QY 791 KRAE-KQKRKMEAKNKENGKVENGLGKTDKRVIVKPEPQVDTEADNMISAVKSR--- 845
DB 1404 KAKEIDNTRSELEKVSLSNDEL-----LEEKQVIK-----SLQDILSYKDKITRN 1450
QY 846 --RLIAIQAKEREITQE-----REMKVKLERQAB--EERIRKHAAAFAFQEGIAK 893
DB 1451 DEKLLSIEDNRKRDLESLEKQRAAQESKAKVEEGLKLEESSSEKAELE-----K 1502
QY 894 AKLVNRRPTIGTDNRNRYWLFSDVPGLFIEKGWVHSDIDYRFNHHCKDHTVSGDEDYC 953
DB 1503 SKEMMKLESTIESNETE-----LKSSME-----TIRKSDKEL 1535
QY 954 PRSKANLGKNAAMTQHGTAATEVAVETTPKQGNLWFLCDSOKELDELLNCLHPQIGIR 1013
DB 1536 ESKKS--AEEDIKNLQH-----BKSLISRINESEKDIIEBKSLRIEAKS 1580
QY 1014 ESQKERLEKRYQDIIHSIHLARKPNLGLKSCDGNQELNLFRLSDILIEVATRLQKGLGY 1073
DB 1581 GSEL-ETVKQLNNAQEKIRINAEENTVLKS-----KLEDIE-----1616
QY 1074 VEETSEFARVLSLEKLDKDFGECVIALQASVTKKFLQGMFAPKQKRLQSEDSAKTEBV 1133
DB 1617 -RELKDKQAEIKNSQEEKE-----LLTSRLKELEQELDSTQQAQKSEERRAEVRKF 1668
QY 1134 DEEKWVREAKVASALE-KWKTAIRAQTFSR-----MHVLLGMLDACIK 1177
DB 1669 QVEKSQLDEK--AMLETKYNDLVNKEQAWKQDEDTVKTTTSQROEIEKLAKELD- 1722
QY 1178 WMSAENARCKVC--PKGEDDKLILCDECNKAFHLFLCLRPALYEPVPGDWQCPACQPAT 1235
DB 1723 -NLKAENSKLKEANEDRSEIDMLMLLVTDLDE-----1753
QY 1236 ARNRRGRNYTESASESDESEDESEDESEDESEDESEDESEDESEDESEDESEDESE 1270
DB 1754 --KNAKYRSKLDLGVIEISSDEEDDEDEDEDEDEDEDEDEDEDEDEDEDEDEDE 1786

RESULT 36
US-10-408-765A-1215
Sequence 1215, Application US/10408765A
Publication No. US20040101874A1
GENERAL INFORMATION:
APPLICANT: Ghosh, Soumitra S.
APPLICANT: Fahy, Bojin D.
APPLICANT: Zhang, Bing
APPLICANT: Gibson, Bradford W.
APPLICANT: Taylor, Steven W.
APPLICANT: Glenn, Gary M.
APPLICANT: Wartock, Dale E.
TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
FILE REFERENCE: 660088.465
CURRENT APPLICATION NUMBER: US/10/408,765A
CURRENT FILING DATE: 2003-04-04
NUMBER OF SEQ ID NOS: 3077
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1215
LENGTH: 1327
TYPE: PRT
ORGANISM: Homo sapiens
US-10-408-765A-1215

Query Match 3.8%; Score 305.5; DB 16; Length 1327;
Best Local Similarity 21.1%; Pred. No. 1.4e-08;

Db 25 PLLDLPPELPERVLQAPPDVGNGEVPKELVHLKLMRKIGSVTADRWEKYLKY 84
QY 46 SRIWTCSTGSSQLTHKEAWEEBOEVAELL-----KEEPPAWTEKLVLEVMVHHTASL 99
Db 85 -----LCECFQDNLKFKNIINEEDADTMRLOPIGRDKGLMYWYQ---LDQDHNVRMYI 136
QY 100 EKLVD---TAWLEIMTKYAVGECDFEVGKERMLVKVIKIHPLEKVEDEATEKSDGAC 156
Db 137 ESQDDQDGSW-----KCIVRNNEIAETLALLKAO-IDPVLLKNSQQDNSSR 184
QY 157 DGPSSDKENSSQIAOHQKKTWVKEDEGRRESINDRARRSPKLTSLKGERKWAPPK 216
Db 185 ESPSLEDEET-----KKEBETPKQEBQKESERKMSBEOQMDL----- 221
QY 217 FLPHKYVVKLQNEDKIISNVPADSLIRTERPPNKIEIVRYFIHNNALRAGTGENAPWVVED 276
Db 222 -----ENRSTANVLSETTVKKEKEDEKELV-----LPVIVKL 254
QY 277 ELVKYSLPSKESDFLLDPYKMTLNPSTKRKNTGSPDKPKSKKSTONSSLSPLNPKL 336
Db 255 E-----KPUPE-----NEEKIIKEESDSFKENVKPIKVEVKECRADPKOTKS----- 297
QY 337 WCHVHLKKSLSPLKVNKSNKS-----PEEHLEEMKMMKSPN-----KLHTN 381
Db 298 ----SWEKPVAGPEPERIEFGGNIKSHETIKSTETETKLNDDQQAQIPLKKEIKLSDD 353
QY 382 FHPKKGPP-----PAKPGKHSK-----PLKAKG-RSKGILNGQKS-----TCN 420
Db 354 PDSVPKGLCKSVTPTEKPLDEIKQEEETCKRISTITGALHEGKQLVNGEVSDERVAPN 413
QY 421 SKSPKGLTKPTK-----MKOMILLDMAKTOQMTAPRNSGGTPTTSKPKHLPAP 474
Db 414 FKTEPIETKFYETKEESYSPSKDRMIITEGNGTESL-----NSVITSMKTGSELEKETAPL 468
QY 475 ALHLIAYYKENDKRDSALSCL--VISTARTLLSSEDRARLPBELSLVQKRYELLEHK 532
Db 469 -----RKDAUSSIVLEHSQAKIEEPD-----PPEMETSLDSS-----EWA 506
QY 533 KWAWMSBEPQRYEYKKGREELKKLKEKAKERREK--EMLRRLKQKRYEDQELTGKNL 590
Db 507 KDLSSKTALSSTESCTMKGEESPKTK---KDKRPPILECLEKESKK-----TFLDK 557
QY 591 PAFRLVDTPEGIP-NTLFGDVAWVFEFLSCYSGLLLPDAQYPIITAVSLMEALSADKGGFL 649
Db 558 DAQRLSPPEEYKSTLESEKPGSP-----AAETSPSPNIIDHCEKLAKE----- 604
QY 650 YLNRVLVILLQTLQDEIAEDYCELGKMLSEIPL-TLHSVSELVRLCLRRSDVQES--- 705
Db 605 -----EVVEQOSTVVG--GQSVKKVDLETLKEDSEFTKVMENDLNDAQTSIE 651
QY 706 EGSDDTDNKSAAFDENVEQDFLEKLETSEFFELTSEBKLIQLTALCHRIILMTYSVQDH 765
Db 652 EPSETKGMQSKFKYKLVPE--BETTASENTEITSERQKE-----GIKLTIRISSR 701
QY 766 METROQMSAELKERLAVLKEENDKKAQKQKKEAKKENGKVENGLGHTDKKRIIV 825
Db 702 KKKPDSPP-----KVLPEKNQKTEKEBEK-----TNVGETLRR----- 736
QY 826 KFEPOVDTBAEDMI SAVKSRRLLAQAKKERIIOERMKVKLERQAEERIKKHAABK 885
Db 737 --SPRI-SRPTAKVAIRDQK--ADKKRGEDEVEEESTALQTKTDKKEILKK-----SEK 787
QY 886 AFQEGIAKAVVRRTPITG-----DRNNHRYWLFPS--DEVPLGFIEKGWVHDSIDYRPN 938
Db 788 DTNSKSVKVK-----PKGVRWVTSRTRGR-WKYSSNDESESGSGSEKS--SAASEEBEE 838
QY 939 HCKDHTVSGDEDYCPRSKANKLGNA-----SWNTOHGTATEVAVETTTTPOGQNLW 991
Db 839 KESEAILADDDDEPC---KKGLPNHPHILILCDSCDSGYHTACLRPLMIIP-DGE--W 892
QY 992 F-----LCS-OKELDELNCLHPQGIRESQKLEKRYQDIH-----SHLAR 1036
Db 893 FCPPCQKLLCEKLEQLQDLVAL-----KKKRAERKERLVVYGISIEIIPQ 944

QY 1037 KPNLG-----LKSCDGNQELNLFRLSLLIEVATRLQK----- 1068
Db 945 EPDFSEDOEEKKOSKSKSKANLLE-----RRSTRTRKCKISYRDFDEFEIDAIEDODIKEA 1000
QY 1069 --GGLYVEETSEFEARVISLEKLDKFCECVIALQASVIKKFLQCFMAPKQKRRKLQSED 1126
Db 1001 DGGVGRGKDI-----TITGHRGKDI-STILDEERKENKRPQRAAAARRKRRRLNDLD 1054
QY 1127 SAKTEVDEKKMVEEAKVASALEKWKTAIREAOTFSRMHVLGLMDACIKWDMGAENAR 1186
Db 1055 S--DSNLDSEES-EDFPKISD-----GSQDEFV--VSDEN-- 1084
QY 1187 KVCPCPKGEDKLLILCDE-----CNKAFHLCFLRPALYVEPDGEWQCPACQATARNRSG 1242
Db 1085 ----PDESEPPPSNDSDTDFCSRRLRRHPSRP-----MQSRRL 1121
QY 1243 RNYTEESASESDEDESDEEEEEEEEDYEVAGLRLRPKRTIRGKHSVI----- 1295
Db 1122 RKPTPKKYSDDEEESEENSRSESDSDFDDFSDDFVETRRRRSRKQKQINYKEDSE 1181
QY 1296 PPAARSGRRPK--KPHSTRRSQPKAPP---VDDAEVDELVLQTKRSRRROSLEQKCE 1349
Db 1182 SDGSKSLRRGKTEIRRVHKKRLSSSESESYLSKNSDEDELAKESKRSVRKGRSTDEYS 1241
QY 1350 EI-----LHKI-----VKYRFSWP-----PREP-- 1367
Db 1242 EADEEEEEEGKPKRKLHRIETDEEESCDNAHGDAQPARDSQPRVLPSEQUESTKKPYR 1301
QY 1368 VTRDEADYDV--ITHPMDFTQV 1389
Db 1302 IESDEEDFENVKVGSLDYSLV 1325

RESULT 39

US-10-094-466-30
; Sequence 30, Application US/10094466
; Publication No. US20030203363A1
; GENERAL INFORMATION:
; APPLICANT: Spytek et al.
; TITLE OF INVENTION: NOVEL HUMAN PROTEINS, POLYNUCLEOTIDES ENCODING THEM
; TITLE OF INVENTION: AND METHODS OF USING
; TITLE OF INVENTION: THE SAME
; FILE REFERENCE: 21402-290D
; CURRENT APPLICATION NUMBER: US/10/094,466
; CURRENT FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/274,281
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/288,148
; PRIOR FILING DATE: 2001-05-02
; PRIOR APPLICATION NUMBER: 60/274,849
; PRIOR FILING DATE: 2001-03-09
; PRIOR APPLICATION NUMBER: 60/275,235
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: 60/338,375
; PRIOR FILING DATE: 2001-12-04
; PRIOR APPLICATION NUMBER: 60/275,579
; PRIOR FILING DATE: 2001-03-13
; PRIOR APPLICATION NUMBER: 60/335,302
; PRIOR FILING DATE: 2001-10-31
; PRIOR APPLICATION NUMBER: 60/275,601
; PRIOR FILING DATE: 2001-03-13
; PRIOR APPLICATION NUMBER: 60/276,000
; PRIOR FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: 60/277,338
; PRIOR FILING DATE: 2001-03-20
; Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: Patin 2.1
; SEQ ID NO 30
; LENGTH: 1398
; TYPE: PRT
; ORGANISM: Homo sapiens

US-10-094-466-30

```
Query Match      3.7%; Score 296; DB 15; Length 1398;
Best Local Similarity 19.2%; Pred. No. 5e-08;
Matches 302; Conservative 249; Mismatches 531; Indels 488; Gaps 78;

QY 42 LERYSERI-----WTCKSTGSSQLTHKEAWEEQVAAELKBEFPANWEKLVLEMY 92
   ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 20 LERYGPDLLPELPSLSW-----TGVQAPRRNGRNGEVPKELVEL-----HLKMRKIG 67
   ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 93 HHTASLEKLVDTAWLEIMTKYAVGECDFE-----VGKEK--- 128
   ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 68 KSVTAD-----RWKYLKILYC--ECQFDDNLKFKNIINEEDATMLQPTIGRDKGL 118
   ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 129 MLKVKTIKVIHPEK-VDEBATEKSDGACDPSDDK-----ENSSQIA 170
   ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 119 MYWQLDQDHNVRMYIEEDQDQSSWKCIVRNRELAETLALLKAQIDPVLLKNSQ-- 176
   ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 171 QHQKQKTVKDEGRRESINDARRSPKLPSTLKKGERKWAPPKFLPHKYDVKLQNEED 230
   ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 177 QDNSSRESPLEDEETKKG-EETPKQEQEKESEKMKSEEQ-----PMDLEN 221
   ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 231 KIISNPADSLIRTPPPNKEIYRIFRHNALRAGTGENAPWVVEDELVKYSLPSKFS 290
   ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 222 RSTANVLEETTVKKEKDEKELVK-----LPVIVKLE--XPLPE----- 258
   ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 291 FLDDPYKYMTLNPSTKRKNTGSPDRKPSKSKTDNSLSPLNPKLWCHVHLKKSLSGSP 350
   ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 259 ---NEEKIIKEESDSFKNVPIKVEKCEKADPKDTKS-----SMEKPPVAQEP 305
   ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 351 LKVNKSNKKS-----PEEHLEEMKMSPN-----KLHTNFHPPKXGP----- 389
   ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 306 ERIEFGNIKSHSHEITEKSTETEKLNDQQAQKIPLKREIKLSDDFSPVKGPLCKSVT 365
   ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 390 PAKKPKKSHDK-----PLKAGK-RSKGLNGOKS-----TGNKSPKPKGLTPKTK 434
   ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 366 PTFEFLKDEIKOBEETCKRISTITLALGHEGKQLVNGEVSDEVRVAPNFKTEPIETKPYETK 425
   ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 435 -----MKQMTLLDMAKGTQKTRAPRNSGGTPRTSSKPKHGLPAPAAHLIAYYKENKDR 488
   ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 426 EESYSPSKDNIITEGNGTESL-----NSVITSMKTGELEKETAPL-----R 467
   ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 489 EKRSALSC--VISTKARLLSSSDRARLPEELSLRSVKRYELLEHKRWASMSSEORKEY 546
   ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 468 KOADSSISVLEIHSQAKIEEPD-----PPMETSLDSS-----EMAKDLSSKTALSSSTES 518
   ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 547 LKKREELKLUKEKAKERREK--EMLERLEKQRYEDQELTGKNIAPFLVDTPEGLP- 603
   ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 519 CTWKGEKSPKTK---KDKRPPILECLEKSKK-----TFLDKDAQRLSPPIPEEVPK 569
   ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 604 NTLFGDVAMVVEFLSCYSGLLLPDAQYPIITAVSLMEALSADKGGFLYLNRLVILLQTL 663
   ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 570 STLESEKPSPE-----AAETSPPNIDHCEKLAKEK-----EVV 605
   ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 664 QDEIADYEGELGKMLSEIPL-TLHSVSELVRLCLRRSDVQEEES---EGSDTDDNKDSAAF 719
   ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 606 ECGSTSTVG--GOSVKVVDLETLEDSEFTKVMNDLNAQTSIGIEPSETKSGMQSKF 663
   ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 720 EDNEVQDEPLEKLETSEFELTSEELQLIITALCHRLIMTYSVQDHMETRQOMSLELWKE 779
   ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 664 KYKLVEP---EETTASENTEITSERQK-----GIKLTIRISSRKKKPSPP----- 707
   ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 780 RLAVLKEENDKRAEKQKMEAKNKGKVENGLGKTDRKRIKVPFEPQVDTAEADMI 839
   ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 708 --KYLEPENKQKTEKEBEK-----TNVGRTLRR-----SPRI-SRPTAKV 745
   ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 840 SAVKSRRLAIQAKKEREIQEREMKVKLRQAEERIRKHAAAEKAFQEGIAKAKLVNR 899
   ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 746 AEIRDQK--ADKRGGEDEVEEESTALQTKDKKELKK---SEKDTNSKVSVKV----- 795
   ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 900 RTPIGT-----DRNHNRYLFS--DEVPLGFTIEKGHWHSIDYRNFHCKHTVSGDEY 952
   ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 796 ---PKGVRWTGRTRGR-WKYSNDESEGSGSEKS---SAASEEKEKESEEAILEDDEP 850
   ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

RESULT 40

```
US-10-408-765A-277
; Sequence 277, Application US/10408765A
; Publication No. US20040101874A1
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fahy, Eoin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; APPLICANT: Warnock, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; FILE REFERENCE: 660088.465
; CURRENT APPLICATION NUMBER: US/10/408,765A
; CURRENT FILING DATE: 2003-04-04
; NUMBER OF SEQ ID NOS: 3077
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 277
; LENGTH: 2375
; TYPE: PRK
; ORGANISM: Homo sapiens
US-10-408-765A-277
```

```
Query Match      3.7%; Score 296; DB 16; Length 2375;
Best Local Similarity 18.5%; Pred. No. 9.6e-08;
Matches 310; Conservative 238; Mismatches 597; Indels 530; Gaps 67;

QY 46 SRIWTCKSTGSGSQTHKEAWEEQVAAELKBEFPANWEKLVLEMYHNTA 97
   ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

Db 188 SNKVVHTSRFPSPKKTSSNCNGEKKLDDSCSGSVTYSALIVPKEMIKAKKLIETTA 247
QY 98 S-----LEKLVDTAWLEIWK-----YAVGECDFV-GKEKWL 130
Db 248 NMNSSVVKFLQATDNSEISSATKLRLQKAFKSVLADIKAHLALEEDLNSEFRAMDAVN 307
QY 131 KYKIVKIHPLEKVDSEATEKSDGACD-----SPSSDKENSQIAQDHQKKEVTVKDE 184
Db 308 KEKNTKEHKVIDAKFETKAKGEKPCALEKKDISKEAKLSKQVDSHEM-HQNYPTBEQ 366
QY 185 GRESINDRARSAPKPLTSLKKGRKWP-----PKFLPHKYDVKLQNEED 230
Db 367 RINKSTGGEHKSDR-----KEPQYEPANTSDELDMDIVSVSPSSVPEDIFENLETAM 419
QY 231 KIISNV--PADSLIRTERPPNKEIVRYFIRHNALRAGTGENAPWVVEDELVKK-----YSLP 285
Db 420 EYQSSVDHGGDSGSGTEQVESSSVKLNISNDNRGGIKSKTAKTVKELYKLPVSLP 479
QY 286 S----KFSDFL-----LDPVKYMTLNPSTKTKGTGSPD-----RK 316
Db 480 NSPIKGADQEVPOKDGQKGLNPKLEKCGLQGENSDNEHLVNEVSLLEESDLRRS 539
QY 317 PSKKS-----KTDNSLSPLNPKLWCHVHLK-KSLSGSPLKVNKSNKSPPEHLEEM 370
Db 540 PRVKTPLRPPTETNPVS--NSDEECNETVKEKQLSVPRKKDKRNS-----DSAI 591
QY 371 KMSPNKLTNPHIPIK-KGPPAKPGKHSKDPKAKGRSKGILNOKSTGNSKSPKGLK 429
Db 592 DNPENKLV-----PKSQSETVDQNSDSDEML-----AILKGVSRMSSSSSDTDIN 638
QY 430 TPTKMKQMTLDMAGTKQMTAPRNSGCTPRTSSKPHKLPPAALHLIAYKENKORE 489
Db 639 EIHTNHK--TLYDL-----KTQAGKDKGKKRKSSTS-----GSDFDTK 676
QY 490 DKRSALSCVISITARLLSEDR---ARLPEELRSL-----VOKRYELLEHKRWASM 538
Db 677 KKSASKSIISKKKQTSSESNYDSELEKEIKSMKIGARTTKR---IPNTKDPSS 733
QY 539 SEQKRYLKKREBELKKLKEKAKERREKEMLEKQKRYEDQELT-GKNLPAPRLVD 597
Db 734 EDEKH-----SKGMDNQGHKLTKEQSSDDAERKQERETSSAEGTVDKDTTLMELRD 789
QY 598 TPEGLPNTLPGDVAMVVEFLSCYSGLLLPDAQYPITANVSLMEALSADKGGFLNRLVI 657
Db 790 R-----LPKKQOASASTDGVDKLSGKQSF-----814
QY 658 LLQTLQDEIADYGE-LGMKLSIPLTLHSVSELVRLCLRRSDVQESSEGSDDTDNKS 716
Db 815 --TSLEVRKVAETKESKHLTKTKCKVQDGLSDIAEKFLKQDSDETSE----DDKQS 868
QY 717 -AAFEDNEVQDFLEK-LETSFFELTSE--EKQLILTALCHRLMTYSVQDHMETROQM 772
Db 869 KKGTEKKKPSDFKKVKIWEQOYESSSDTEKLPEREICHFPKGIKQKNGTTDGEKK 928
QY 773 SAEKWERLAVLKEENDKR-----AEQKKEMEAKNKGKVEG-----L 815
Db 929 SKK-----RDTKSKKKBELSDYAEKSTGKSDSCSDSKSNGAYGKRCCKLL 980
QY 816 GKTDRKRIKVPQPDTE-----AEDMISAVSRRLAIQAKKER-----BIQERE 862
Db 981 GASSRKQDCS---SSDTKYSMKEDGCSN-SDKRLKRIELRERNLSSKRNKTKIQSGS 1036
QY 863 MKVKLERQAEERIRKHKAASKAQOEGIAKALV--MRPTPIGTDRNNHNRVWLFSD--- 917
Db 1037 SSSDAEESSEDNKKKQRTSSKK-----KAVIVKEKRNLSRTSTKRQADITSSSSS 1089
QY 918 -----EVPGLFIEKGWHDSDYRFNHCHCKDTVSGDEYVC 953
Db 1090 DIEDDDQNSIGSGSDEQIKIPVTENLVLSHTGFCQSGGDEALSKSPVTVDDDDDDND 1149
QY 954 PRSK-----KANLGKNASMNTQHGTAETAVETTTTPQGNLWFLCDSQKLEOEL 1003
Db 1150 PENRIAKKMLLEIKANLSSD-----EDGSSDDEPEEGKRTGKN-----HENPGDEEA 1199

QY 1004 LNCILHPQIRSQERLEKRYQDIHIIHILARKPNLGLKSCDGNQELLNFLRSLIEVA 1063
Db 1200 KNQVNSDSDESSESCKPRYRHLRLRHLKLTVS-----DGES-----1235
QY 1064 TRLQGGIGYVEETSEPEARVISLEKLDQFGEVIALQASVIKFLQGFMAPKQKRRKLQ 1123
Db 1236 -----GEBKTKPKHEKVK-----GRNRRKVS 1258
QY 1124 SED-----SAKTEEV-----DEEKQVBEAKVASALEKWKTAIREAQTFSRMHVLLGM 1171
Db 1259 SEDSESDSDFQSSGVSEVSESEDEQRPRTSRAKAE-LEENQRSYKQKKRRRIKV----1313
QY 1172 LDACIKWDSABENARCKVCPKKGEDDKLILCDECNKAFHLFCLRPALYEPDGEWQCPC 1231
Db 1314 -----QEDSSSEN-----1321
QY 1232 QPATARRNSRGRNYTEESASEDSDESDSEDEEBEEEEEBEEDYEAGLRLRPRTIRGK 1291
Db 1322 -----KSNSEEEEEKEEEEEEEEEEEED-----END 1351
QY 1292 HSVIPPAARSRRRPGKPHSTRSQPKAPPVDDAEVDELVLQTKRSSRRSQLELQKCEI 1351
Db 1352 DSKSPGKGRKIRKILKDDKLRTEQNA-----LKEEBERRKRIAEREREREK 1399
QY 1352 LHKIVKYRFPSPFPVTR-----DEAEDYDVIT-----HPMDFTVQNKCS 1395
Db 1400 LREVIIEIDASPTKCPITTKLVLDDEDETKPLVQVHRNMVILKPHQVQGVQFWMDCCC 1459
QY 1396 GYSRSVQBFITDMKQVFTNAEVYNCRGSHVLSVMVTEQCLVLLHKHL-----1444
Db 1460 ESVKTK-----KSPGSGCILAHCWG-----LGTILQ--VVSFLATVLLCDKLDFTAL 1506
QY 1445 ---FGHPYVR--RRKKFPDLAEDGDESEPEAVGQSRDEDRRSREAEIQEWLQD 1494
Db 1507 VVCPLANTALNMNBEKQWEGKDKDEKLEVSELATVTRPQE---RSYMLQWQED 1558

RESULT 41

US-10-369-493-5368
; Sequence 5368, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:

; APPLICANT: Cao, Yongwei

; APPLICANT: Hinkle, Gregory J.

; APPLICANT: Slater, Steven C.

; APPLICANT: Goldman, Barry S.

; APPLICANT: Chen, Xianfeng

; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

; FILE REFERENCE: 38-10(52052)B

; CURRENT APPLICATION NUMBER: US/10/369,493

; CURRENT FILING DATE: 2003-02-28

; PRIOR APPLICATION NUMBER: US 60/360,039

; PRIOR FILING DATE: 2002-02-21

; NUMBER OF SEQ ID NOS: 47374

; SEQ ID NO 5368

; LENGTH: 1827

; TYPE: PRT

; ORGANISM: Caenorhabditis elegans

Query Match 3.7%; Score 295; DB 15; Length 1827;

Best Local Similarity 20.2%; Pred. No. 7.9e-08;

Matches 317; Conservative 261; Mismatches 627; Indels 366; Gaps 72;

QY 121 DPEVCKEM-----LKVKIVKIHPLEKVDSEATEKSDGACDSPSS---DKEN--SSOIA 170

Db 55 DIGVGENLDELFAKLEELFNNTLSEVNEMLRE-ENDAALAANEHLRVDATNLSRQ 113

QY 171 QDHQKKEVTVKDEGRRESINDRARSRKLP-----TSKKGERKWWP 215

Db 114 QLQOQOHT-----ESMRFRSNTSRELQYCKGPEICNTFLFLFSCPKPYQERYRNQ 164


```
QY 503 ARLSSDRARPEELRSLVKRYELHEHKKRWASMSBEQRKEYLKKKREELKKLKEKA 562
Db 1226 KOTLEKEN-ADLAGELRVLGQAKQB-VEHKKK-----KLEAQVQLQSKCSD-- 1270
QY 563 KERREKEMLERLEKOKRYEDQELTGKNLPAPFLVDTPEGLPNTLFGDVAMVVFSLCYSG 622
Db 1271 GERARAEIENDKVHKLQN-EVESVTG-----MLNEAEGKAIKLAKDVASLSSQLQDTQE 1322
QY 623 LLLPDAQYPTAVSLMEALSADKGGFLYNRVLVILLQTLLODETAED-----YG 672
Db 1323 LLOEETROKLVNSTKLRLQLEERNS-----LQDLDDEMEAKQNLERHIS 1367
QY 673 ELGKMLSEIPLTLHVSSELVR-LCLRRSDVQSESGSDTDNKSAAFE-----DNEVQD 726
Db 1368 TLNIQLSDSKKQLQDFASTVEALEBGGKRFQKEIENLTQOYEEKAAAYDKLEKTKNRLQ 1427
QY 727 EF-----LEKLETSEFFELTSEEKQLILTALCHRLMTYSVQDHMETRQ 770
Db 1428 ELDDLVDLDNQRLVSNLEK-KORKFDQLAEEK-----NISSKYA-----DERD 1472
QY 771 QMSAEL-WKERLAV-----LKEENDKKRAEKOKRKEKEMAKNKENGKVENGLGKTRD 820
Db 1473 RAEAREKETHKALSLARALEALEAKELELERTNMLKAEMEDLVSSKDDVGNVHELEK 1532
QY 821 KKRIVKFEPOVD-----TEADMTSAVKSRL-----LAIQAKEREIOER-----EM 863
Db 1533 SKRAL-ETQMBEMTKQLELEDELOATEDAKRLLEVNNQALKGQFERDLOARDQNEBK 1590
QY 864 KVKLERQ-----AEEERIRKHAAAEKAFQEGIAKVLNVRTPI-GTDNRHNRMYLF 915
Db 1591 RQQLRQLHEYTELEDERKORALAAAKKLEGLDKOLELQADSAIKGRE-----1641
QY 916 SDEVFGLPTEKGVWHDSDYRNNHCKDHTVSGDDYC---PRSKANIGKNVNASMNTQHG 972
Db 1642 -BAIKQLRKLQAKMD-----FORELEDAARSDEIFATAKENEKKAKSLEADLMQLOED 1695
QY 973 TAT-----EVAVETTPKQGNLWFLCDSQKELDELNCLNHPQGIRESOLKE 1019
Db 1696 LAAERARKQADLEKEELAEALASSLSGRNA--LQDEKKRLB-----ARTAQLEE 1743
QY 1020 RLEKRYQDIHSHILARKPNGLKSCDGNQELNLFRLSDLIBEAVTRLQKGLGYVEETSE 1079
Db 1744 ELEEE-----QGNWEAM---SDRVKATQ-----1764
QY 1080 FEARVISLEKLDGFCVJALQASVIKKFLOGMAPKQKRLQSEDSAKTEVDDEKKN 1139
Db 1765 -QAEQLSNE-----LATERSTAQK-----NESARQQLERQNKELRSLKHEMEGA 1807
QY 1140 VE-----BAKVASALEKWKTAIREAQTFSRMHVLLGMLDACIKWDMGAENARCKV 1189
Db 1808 VSKPKFSTTAALEAKIAQLEEQVEAREKQAATK-----1842
QY 1190 CPKKGEDDKLILDCENKAFHFLCLRPALEYVDPGEWQCPAQCPATARNNSGRNYTE-- 1247
Db 1843 -SLKQKDK-----LKEILLQVEDERKMAEQYKEQAEGNARVQLKRLQ 1886
QY 1248 ESASDSEDSDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDE 1304
Db 1887 EBAEESORINANRRKLQELDEATESENEAMGREVNALKSLRRGNETSFVPSRRSGGR 1945
```

RESULT 45

```
US-10-341-434-103
; Sequence 103, Application US/10341434
; Publication No. US20030215835A1
; GENERAL INFORMATION:
; APPLICANT: OriGene Technologies
; TITLE OF INVENTION: Differentially Regulated Prostate Cancer Genes
; FILE REFERENCE: 9U 204 205 R1
; CURRENT APPLICATION NUMBER: US/10/341,434
; CURRENT FILING DATE: 2003-07-18
; PRIOR APPLICATION NUMBER: US 60/348,164
; PRIOR FILING DATE: 2002-01-15
```

```
; PRIOR APPLICATION NUMBER: US 60/348,119
; PRIOR FILING DATE: 2002-01-15
; NUMBER OF SEQ ID NOS: 238
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 103
; LENGTH: 1972
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-341-434-103
```

```
. Query Match 3.7%; Score 292.5; DB 15; Length 1972;
Best Local Similarity 19.9%; Pred. No. 1.2e-07;
Matches 275; Conservative 214; Mismatches 507; Indels 383; Gaps 58;
```

```
QY 75 LKBEFPWYKEL-VLEWHHTASLEKLVDTAWLEIMTKY-----AVGERCD 121
Db 801 LARKAFARQOQLTAMKVIQNRCAAYLKLKNQWMLFTKVKPLQLQVTRQBEEMQAKDE 860
QY 122 FEVGKEKMLVKIVIHLEKVDDEATEKKS-----DGACDSPSSDKENSQIAQDHQK 175
Db 861 LQTKEROQKAB-NELKELEBQKHSQLETEKLLQLOLAETELYAEAEEMRVRLAAKQOE 919
QY 176 KETVVEDEGRRESINDRARESPRKLPTSLKKGERKWAPPKPLPHKYDVKLQNEDKIISN 235
Db 920 LEEILHEMARLEBEEDRGQ-----LQAEK-----KWAQQLDLEEQLER- 961
QY 236 VPADSLIATERPPNKEIVYRIRHVALRAGTGENAPWVVEDELVKKYSLPSPKFS---DPL 292
Db 962 -----EAAARQKLQ---LEKVTAETAKIKKLEDEILYMDQNNKLSKERKLL 1004
QY 293 LDPYKYMTLNSTKTKNGTSPDRPSKSKTDNSSLSPNPLNPKWCHVHLKKSLSG--- 348
Db 1005 EERISDLTTLNABEEKA-----KULTKLNKHESMISEL-----EVLKKEEKSRLQEL 1053
QY 349 SPLKYNSKNSKSPBEHLEEMKMMSPNKLHTNPHIPKGGPPAKPGKXHSKPLKAKGRS 408
Db 1054 EKLKEKLGSDADPHEQIADLQAQIAELKM---QLANKEEBELQALARLDDEIAQKNA 1109
QY 409 -KGI-----LNGQKSTGNSKSPK-----GLKTPKTQKQMTLLDMAKGTQK 449
Db 1110 LKKIRELBGHISDLOEDLDSERARNKAEBKQKRDIGEELKTELED--TLD-STATQ 1166
QY 450 MTRAPNRSGGT-----PRTSSKPKHLPAPAAHLIAVYKENKDKRSALSVCISKT 502
Db 1167 ELRAKREBEVTVLKALDEETSHQAQVQEMRQKHAQAVEELTEQLEQFKRA-KANLDKN 1225
QY 503 ARLSSDRARPEELRSLVKRYELHEHKKRWASMSBEQRKEYLKKKREELKKLKEKA 562
Db 1226 KOTLEKEN-ADLAGELRVLGQAKQB-VEHKKK-----KLEAQVQLQSKCSD-- 1270
QY 563 KERREKEMLERLEKOKRYEDQELTGKNLPAPFLVDTPEGLPNTLFGDVAMVVFSLCYSG 622
Db 1271 GERARAEIENDKVHKLQN-EVESVTG-----MLNEAEGKAIKLAKDVASLSSQLQDTQE 1322
QY 623 LLLPDAQYPTAVSLMEALSADKGGFLYNRVLVILLQTLLODETAED-----YG 672
Db 1323 LLOEETROKLVNSTKLRLQLEERNS-----LQDLDDEMEAKQNLERHIS 1367
QY 673 ELGKMLSEIPLTLHVSSELVR-LCLRRSDVQSESGSDTDNKSAAFE-----DNEVQD 726
Db 1368 TLNIQLSDSKKQLQDFASTVEALEBGGKRFQKEIENLTQOYEEKAAAYDKLEKTKNRLQ 1427
QY 727 EF-----LEKLETSEFFELTSEEKQLILTALCHRLMTYSVQDHMETRQ 770
Db 1428 ELDDLVDLDNQRLVSNLEK-KORKFDQLAEEK-----NISSKYA-----DERD 1472
QY 771 QMSAEL-WKERLAV-----LKEENDKKRAEKOKRKEKEMAKNKENGKVENGLGKTRD 820
Db 1473 RAEAREKETHKALSLARALEALEAKELELERTNMLKAEMEDLVSSKDDVGNVHELEK 1532
QY 821 KKRIVKFEPOVD-----TEADMTSAVKSRL-----LAIQAKEREIOER-----EM 863
Db 1533 SKRAL-ETQMBEMTKQLELEDELOATEDAKRLLEVNNQALKGQFERDLOARDQNEBK 1590
```


; SEQ ID NO 67
; LENGTH: 65
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-702-148-67

Query Match 3.5%; Score 281; DB 15; Length 65;
Best Local Similarity 92.3%; Pred. No. 7.8e-09;
Matches 60; Conservative 1; Mismatches 0; Indels 4; Gaps 1;

Qy 630 YPITAVSLMEALSADKGGFLYLNRVLVI---LLQTLQLQDEIAEDYCELGKLSKIP LTL 685
Db 1 YPITAVSLMEALSADKGGFLYLNRVLVILLQTLQLQDEIAEDYCELGKLSKIP LTL 60

Qy 686 HSVSE 690
Db 61 HSVSE 65

Search completed: March 9, 2005, 15:10:54
Job time : 168.022 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 9, 2005, 13:06:37 ; Search time 110.673 Seconds
(without alignments)
5336.298 Million cell updates/sec

Title: US-10-702-148-27
Perfect score: 7967
Sequence: 1 MAPLGRKPFPLVNPFGEE.....NCFMMLVNTQFCMALTDVT 1527

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 65 summaries

Database : A_Geneseq_16Dec04:*1: Geneseq1980s:*2: Geneseq1990s:*3: Geneseq2000s:*4: Geneseq2001s:*5: Geneseq2002s:*6: Geneseq2003as:*7: Geneseq2003bs:*8: Geneseq2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	7967	100.0	1527	2 AAW81172	Aaw81172 Human BAZ
2	7955	99.8	1531	2 AAW81173	Aaw81173 Human BAZ
3	7670	96.3	1483	8 ADQ20910	Adq20910 Human sof
4	7668	96.2	1483	7 ADF69140	Adf69140 Human MPS
5	5319	66.8	1052	6 ABR41364	Abr41364 Human DIT
6	2889	36.3	572	4 AAM14838	Aam14838 Peptide #
7	2889	36.3	572	4 ABB33805	Abb33805 Peptide #
8	2889	36.3	572	4 AAM27265	Aam27265 Peptide #
9	2889	36.3	572	4 ABB28622	Abb28622 Peptide #
10	2889	36.3	572	4 ABB19248	Abb19248 Protein #
11	2889	36.3	572	4 AAM66978	Aam66978 Human bon
12	2889	36.3	572	4 AAM54571	Aam54571 Human bra
13	2889	36.3	572	4 ABG48640	Abg48640 Human liv
14	2889	36.3	572	4 AAM02563	Aam02563 Peptide #
15	2889	36.3	572	5 ABB36635	Abb36635 Human pep
16	2847	35.7	560	4 AAM18320	Aam18320 Peptide #
17	2847	35.7	560	4 ABB37354	Abb37354 Peptide #
18	2847	35.7	560	4 AAM30808	Aam30808 Peptide #
19	2847	35.7	560	4 ABB32101	Abb32101 Peptide #
20	2847	35.7	560	4 ABB22640	Abb22640 Protein #
21	2847	35.7	560	4 AAM70484	Aam70484 Human bon
22	2847	35.7	560	4 AAM58044	Aam58044 Human bra
23	2847	35.7	560	4 AAM05928	Aam05928 Peptide #
24	2847	35.7	560	5 ABB40123	Abb40123 Human pep
25	2305.5	28.9	513	4 AAM00760	Aam00760 Human bon

26	1078	13.5	209	6	ABR41346	Abr41346 Human DIT
27	929.5	11.7	1674	2	AAW81169	Aaw81169 Human BAZ
28	929.5	11.7	1674	2	ADF69139	Adf69139 Human MPS
29	929.5	11.7	1674	8	ADP12578	Adp12578 Protein e
30	924.5	11.6	1540	2	AAV07734	Aay07734 Human htl
31	712	8.9	141	4	AAU16181	Aau16181 Human nov
32	712	8.9	141	6	ABU55250	Abu55250 Human nov
33	664	8.3	1476	4	ABBS8706	Abbs8706 Drosophil
34	637	8.0	128	3	AAG03625	Aag03625 Human sec
35	605	7.6	1586	7	ABM85416	Abm85416 Mouse pro
36	540.5	6.8	1873	7	ABM85417	Abm85417 Human pro
37	537.5	6.7	1727	4	AAW85554	Aaw85554 Human pro
38	537.5	6.7	1878	4	AAW40239	Aam40239 Human pol
39	536	6.7	1589	4	AAW42025	Aam42025 Human pol
40	532	6.7	1972	2	AAW81171	Aaw81171 Human BAZ
41	532	6.7	1972	6	ABR64241	Abr64241 Angiogene
42	532	6.7	1972	8	ADP54420	Adp54420 Human PRO
43	523.5	6.6	1878	2	AAW81170	Aaw81170 Human BAZ
44	416.5	5.2	475	4	AAW93746	Aab93746 Human pro
45	368.5	4.6	524	8	ADR09529	Adr09529 Human pro
46	352	4.4	60	3	AAG03100	Aag03100 Human sec
47	347.5	4.4	1264	6	ABR52663	Abr52663 Protein s
48	347.5	4.4	1264	7	ADK61762	Adk61762 Diseasee t
49	344	4.3	2897	4	ABBS8514	Abbs8514 Drosophil
50	341	4.3	65	4	AAW14835	Aam14835 Peptide #
51	341	4.3	65	4	ABR33802	Abb33802 Peptide #
52	341	4.3	65	4	AAW27262	Aam27262 Peptide #
53	341	4.3	65	4	ABR28619	Abb28619 Peptide #
54	341	4.3	65	4	ABR19245	Abb19245 Protein #
55	341	4.3	65	4	AAW66975	Aam66975 Human bon
56	341	4.3	65	4	AAW54568	Aam54568 Human bra
57	341	4.3	65	4	ABG48637	Abg48637 Human liv
58	341	4.3	65	4	AAW02560	Aam02560 Peptide #
59	341	4.3	65	5	ABG36632	Abg36632 Human pep
60	338	4.2	69	4	AAW14836	Aam14836 Peptide #
61	338	4.2	69	4	ABR33803	Abb33803 Peptide #
62	338	4.2	69	4	AAW27263	Aam27263 Peptide #
63	338	4.2	69	4	ABR28620	Abb28620 Peptide #
64	338	4.2	69	4	ABR19246	Abb19246 Protein #
65	338	4.2	69	4	AAW66976	Aam66976 Human bon

ALIGNMENTS

RESULT 1
AAW81172
ID AAW81172 standard; protein; 1527 AA.

XX AAW81172;

XX 05-MAY-1999 (first entry)

DT Human BAZ1-beta protein #1.

DE Transcriptional regulator; transcription; BAZ1-alpha; bromodomain; BAZ;
KW atypical zinc finger; testis; human; tumour; BAZ1-beta; BAZ2-alpha; drug;
KW BAZ2-beta; treatment; cancer; proliferative disorder; screening.

XX Homo sapiens.

OS WO9847920-A1.

PN 29-OCT-1998.

PD 17-APR-1998; 98WO-JP001783.

PF 18-APR-1997; 97JP-00116570.

PR 24-OCT-1997; 97JP-00310027.

XX (CHUS) CHUGAI RES INST MOLECULAR MEDICINE INC.

PA Jones MH;

PI

XX WPI: 1998-583603/49.
DR N-PSDB; AAV68408.
XX
PT Transcriptional regulator gene family containing bromodomain - may be
PT expressed in testis tissue and is useful for treatment of cancer and
PT other proliferative disorders.
XX
PS Claim 1; Page 125-137; 187pp; Japanese.
XX
CC This sequence represents the human BAZ1-beta protein, a member of a
CC family of transcriptional regulator genes containing a bromodomain (BAZ,
CC Bromodomain with Atypical Zinc finger) which are expressed specifically
CC in testis tissue and also in certain tumour lines. Transgenic cells may
CC be used for the preparation of the BAZ1-alpha, BAZ1-beta, BAZ2-alpha and
CC BAZ2-beta proteins. These proteins can be used in the treatment of cancer
CC and other proliferative disorders, and in screening of compounds for
CC their binding ability to the expression products (e.g. for use as drugs
CC by modulation of transcriptional regulation)
XX
SQ Sequence 1527 AA;
Query Match 100.0%; Score 7967; DB 2; Length 1527;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1527; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MAPLGRKPPFLVNPPLGEEPPFTTPIHTQEAFTREYEARELRYSERIWTCKSTGSSQL 60
Db 1 MAPLGRKPPFLVNPPLGEEPPFTTPIHTQEAFTREYEARELRYSERIWTCKSTGSSQL 60
QY 61 THKEAWEEQVAAELKEEFPAPWYKLVLEWVHNTASLEKLVDPFAWLEIMTKYAVGEEC 120
Db 61 THKEAWEEQVAAELKEEFPAPWYKLVLEWVHNTASLEKLVDPFAWLEIMTKYAVGEEC 120
QY 121 DFEVGKERMLKVKIKIHPLEKVDDEATEKSDGACDSPSSDKENSSQIAOHOKKETVV 180
Db 121 DFEVGKERMLKVKIKIHPLEKVDDEATEKSDGACDSPSSDKENSSQIAOHOKKETVV 180
QY 181 KEDEGRRESINDRARRSPKLPSTSLKGERKWAPPKFLPHKYDVVKLQNEDKIISNVPADS 240
Db 181 KEDEGRRESINDRARRSPKLPSTSLKGERKWAPPKFLPHKYDVVKLQNEDKIISNVPADS 240
QY 241 LIRTERPNKEIVRFIERNALRAGTGENAPWVEDELVKYSLPSKESDFLLDPYKYMT 300
Db 241 LIRTERPNKEIVRFIERNALRAGTGENAPWVEDELVKYSLPSKESDFLLDPYKYMT 300
QY 301 LNPSTKRNTGSPDRKPKSKTDSNLSSPLNPKLWCHVHLKXSLGSPKLVKNSKNSK 360
Db 301 LNPSTKRNTGSPDRKPKSKTDSNLSSPLNPKLWCHVHLKXSLGSPKLVKNSKNSK 360
QY 361 SPEEHLEENMKWMSPNKLTNPHIPKGGPPAKPKGHSKDKPLKAGRSKGLNGOKSTGN 420
Db 361 SPEEHLEENMKWMSPNKLTNPHIPKGGPPAKPKGHSKDKPLKAGRSKGLNGOKSTGN 420
QY 421 SKSPKGLKTPKTKWQMTLLDMAKTQMTAPNSGGTPTRTSSKPHKHPALHILIA 480
Db 421 SKSPKGLKTPKTKWQMTLLDMAKTQMTAPNSGGTPTRTSSKPHKHPALHILIA 480
QY 481 YYKENKREDKRSALSCVISKTARLLSSDDRARLPPEELRSIVQKRYELLEHKRWASME 540
Db 481 YYKENKREDKRSALSCVISKTARLLSSDDRARLPPEELRSIVQKRYELLEHKRWASME 540
QY 541 EQRKYLKKREELKKLKEKAKERREKEMLERLEKQKRYEDQELTGKVLPAFRLVDTPPE 600
Db 541 EQRKYLKKREELKKLKEKAKERREKEMLERLEKQKRYEDQELTGKVLPAFRLVDTPPE 600
QY 601 GLPNTLFGDVAMVVEFLSCYSGLLLPDAQYPTAVSLMEALSADKGGFLYNRLVILLQ 660
Db 601 GLPNTLFGDVAMVVEFLSCYSGLLLPDAQYPTAVSLMEALSADKGGFLYNRLVILLQ 660
QY 661 TLLODEIAEDYGELGKMLSEIPLTLHVSSELVRLCLRRSDVQEESEGSDTDNDKDSAAFE 720
Db 661 TLLODEIAEDYGELGKMLSEIPLTLHVSSELVRLCLRRSDVQEESEGSDTDNDKDSAAFE 720

QY 721 DNEVQDFLEKLETSEFPFELTSEBKLIQILTALCHRIILMTYSVQDHMETROQMSAELWKER 780
Db 721 DNEVQDFLEKLETSEFPFELTSEBKLIQILTALCHRIILMTYSVQDHMETROQMSAELWKER 780
QY 781 LAVLKEENDKKRAEKQKRAKEMAKNKGKVENGLGKTDRKKRIVKFPQVDTEADMTIS 840
Db 781 LAVLKEENDKKRAEKQKRAKEMAKNKGKVENGLGKTDRKKRIVKFPQVDTEADMTIS 840
QY 841 AVKSRRLAIQAKKEREIQEREMVKLERQAEERIRKHAAAEKAFQEGIAKAKLVNMR 900
Db 841 AVKSRRLAIQAKKEREIQEREMVKLERQAEERIRKHAAAEKAFQEGIAKAKLVNMR 900
QY 901 TPIGTDRNHNRYWLFSEDEVPLGFTKGVWVHSDIDYRFNHNCKDHTVSDGDEYCPRSKAN 960
Db 901 TPIGTDRNHNRYWLFSEDEVPLGFTKGVWVHSDIDYRFNHNCKDHTVSDGDEYCPRSKAN 960
QY 961 LGKNASMTQGTATVAVETTTTPQGNLWFLCDQSKELDELNCLHPQGIRESQLKER 1020
Db 961 LGKNASMTQGTATVAVETTTTPQGNLWFLCDQSKELDELNCLHPQGIRESQLKER 1020
QY 1021 LEKRYQDIHLSIHLARKPNLGLKSCDGNQELLNFIKSDLIIEVATRLQKGLGYVEETSEF 1080
Db 1021 LEKRYQDIHLSIHLARKPNLGLKSCDGNQELLNFIKSDLIIEVATRLQKGLGYVEETSEF 1080
QY 1081 EARVISLEKLDKDFGECVIALQASVKKFLOGFMAPKQKRRKLOSEDSAKTEVDEEKKWV 1140
Db 1081 EARVISLEKLDKDFGECVIALQASVKKFLOGFMAPKQKRRKLOSEDSAKTEVDEEKKWV 1140
QY 1141 BEAKVASALEKWKTAIREAQTFSRMVLGLMLDACIKWDMSAENARCKVCPKKGDDKLI 1200
Db 1141 BEAKVASALEKWKTAIREAQTFSRMVLGLMLDACIKWDMSAENARCKVCPKKGDDKLI 1200
QY 1201 LCDECNKAFHFLCLRPALEYVDPGEWQCPACQATARRNSGRNYTESASEDSEDDSD 1260
Db 1201 LCDECNKAFHFLCLRPALEYVDPGEWQCPACQATARRNSGRNYTESASEDSEDDSD 1260
QY 1261 EEEEEEEEEDEEYVAGLRLPRKTIKRGHSVIPPAAARSGRRPKKPHSTRSQKAP 1320
Db 1261 EEEEEEEEEDEEYVAGLRLPRKTIKRGHSVIPPAAARSGRRPKKPHSTRSQKAP 1320
QY 1321 PVDDAEVDVLQTKRSRRSLELQKCEELHKIKYKRFWSWFPREPVTREAEYDVI 1380
Db 1321 PVDDAEVDVLQTKRSRRSLELQKCEELHKIKYKRFWSWFPREPVTREAEYDVI 1380
QY 1381 THPMDFQTVQNKSCGYSYRSVQEFLLDMKQVFTNAEVNCRGSHVLSQWVKTEQCLVLL 1440
Db 1381 THPMDFQTVQNKSCGYSYRSVQEFLLDMKQVFTNAEVNCRGSHVLSQWVKTEQCLVLL 1440
QY 1441 HKHLPCHPYVRRKPKPPDLAEDGSEPAVQSRDEDRSREAEIQEWLQDTSLSYA 1500
Db 1441 HKHLPCHPYVRRKPKPPDLAEDGSEPAVQSRDEDRSREAEIQEWLQDTSLSYA 1500
QY 1501 KINSKDHNCFMVLVNTQFCMALTDVT 1527
Db 1501 KINSKDHNCFMVLVNTQFCMALTDVT 1527
RESULT 2
AAW81173
ID AAW81173 standard; protein; 1531 AA.
XX
AC AAW81173;
XX
DT 05-MAY-1999 (first entry)
XX
DE Human BAZ1-beta protein #2.
XX
KW Transcriptional regulator; transcription; BAZ1-alpha; bromodomain; BAZ;
KW atypical zinc finger; testis; human; tumour; BAZ1-beta; BAZ2-alpha; drug;
KW BAZ2-beta; treatment; cancer; proliferative disorder; screening.
OS Homo sapiens.

[illegible]

ID ADQ20910 standard; protein; 1483 AA.
 XX AC ADQ20910;
 XX DT 26-AUG-2004 (first entry)
 XX DE Human soft tissue sarcoma-upregulated protein - SEQ ID 3730.
 XX KW soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human.
 XX OS Homo sapiens.
 XX PN WO2004048938-A2.
 XX PD 10-JUN-2004.
 XX PF 26-NOV-2003; 2003WO-US038193.
 XX PR 26-NOV-2002; 2002US-0429739P.
 XX PA (PROT-) PROTEIN DESIGN LABS INC.
 XX PI Aziz N, Ginsburg WM, Zlotnik A;
 XX WI WIPI; 2004-441208/41.
 XX DR Early detection of soft tissue sarcoma comprises determining expression
 XX PT of a gene in a first soft tissue sample and a normal soft tissue sample
 XX PT and comparing the gene expression, also useful in treating soft tissue
 XX PT sarcoma.
 XX PS Example 2; SEQ ID NO 3730; 210pp; English.
 XX CC The invention relates to a novel method for detecting soft tissue sarcoma
 XX CC which comprises obtaining a first soft tissue sample from an individual
 XX CC and a normal soft tissue sample from the same or different individual,
 XX CC determining the expression of a gene in both samples and comparing the
 XX CC expression of the gene in both soft tissue samples, where a higher level
 XX CC of protein expression in the first soft tissue sample indicates the
 XX CC presence of soft tissue sarcoma. The method of the invention has
 XX CC cytostatic applications and may be useful for detecting soft tissue
 XX CC sarcoma, possibly via gene therapy or vaccine production. The nucleic
 XX CC acid sequences may be useful in diagnostic and screening applications.
 XX CC The current sequence is that of a human soft tissue sarcoma-upregulated
 XX CC protein of the invention. The current sequence is not shown within the
 XX CC specification per se but was submitted in CD format by the inventor.
 XX SQ Sequence 1483 AA;

Query Match 96.3%; Score 7670; DB 8; Length 1483;
 Best Local Similarity 99.2%; Pred. No. 0;
 Matches 1473; Conservative 3; Mismatches 7; Indels 2; Gaps 1;

QY 1 MAPLLGRKPPFLVNPDPGEEPPFTTPTHTQEAFTREEYEARLERYSERIWTCKSGSSQL 60
 DB 1 MAPLLGRKPPFLVNPDPGEEPPFTTPTHTQEAFTREEYEARLERYSERIWTCKSGSSQL 60
 QY 61 THKEAWEEQEVAELLKEEFPWYKLVLEMVHHTASLEKLVDTAWLEIMTKYAVGEEC 120
 DB 61 THKEAWEEQEVAELLKEEFPWYKLVLEMVHHTASLEKLVDTAWLEIMTKYAVGEEC 120
 QY 121 DFEVGEKMKLVKIYIHPLEKVDTEATEKSDGACDSSPKENSSQIAQDHQKETT 180
 DB 121 DFEVGEKMKLVKIYIHPLEKVDTEATEKSDGACDSSPKENSSQIAQDHQKETT 180
 QY 181 KEDEGRRESINDRARRSPKLTSLKGERKWAPPKFLPHKYDVKLQNEDKIISNPADS 240
 DB 181 KEDEGRRESINDRARRSPKLTSLKGERKWAPPKFLPHKYDVKLQNEDKIISNPADS 240
 QY 241 LIRTERPPNKEIVRYFIRHNALRAGTGENAPWVEDELVKKYSLPKSFDFLLDPKYMT 300
 DB 241 LIRTERPPNKEIVRYFIRHNALRAGTGENAPWVEDELVKKYSLPKSFDFLLDPKYMT 300

QY 301 LNPSTKRKNTGSPDRKPSKSKTDNSLSLSSPLNPKLWCHVHLKKLSLSPKLVKNSKNSK 360
 DB 301 LNPSTKRKNTGSPDRKPSKSKTDNSLSLSSPLNPKLWCHVHLKKLSLSPKLVKNSKNSK 360
 QY 361 SPEEHLEEMMOMGPNKLTHTNFHIPKGGPPAKPGKHSDKPLKAKGRSGKILNGQKSTGN 420
 DB 361 SPEEHLEEMMOMGPNKLTHTNFHIPKGGPPAKPGKHSDKPLKAKGRSGKILNGQKSTGN 420
 QY 421 SKSPKKGKLTPTKTKMKTLDMAKGTQKMTAPRNSGGTPTRTSSKPKHKLPPAALHLIA 480
 DB 421 SKSPKKGKLTPTKTKMKTLDMAKGTQKMTAPRNSGGTPTRTSSKPKHKLPPAALHLIA 480
 QY 481 YYKENKREDKRSALSVCVISTARLLSSEDRARLPEELRSLVQKRYELLEHKRWASMS 540
 DB 481 YYKENKREDKRSALSVCVISTARLLSSEDRARLPEELRSLVQKRYELLEHKRWASMS 540
 QY 541 EORXEYLKKRBEELKKLKEKAKERREKEMLERLEKQRYEDQELTKNLPFRVLDTPE 600
 DB 541 EORXEYLKKRBEELKKLKEKAKERREKEMLERLEKQRYEDQELTKNLPFRVLDTPE 600
 QY 601 GLPNTLFGDVAMVVEFLSCYSGLLLPDAQYPTITAVSLMEALSADKGGFLYLNRLVILLQ 660
 DB 601 GLPNTLFGDVAMVVEFLSCYSGLLLPDAQYPTITAVSLMEALSADKGGFLYLNRLVILLQ 660
 QY 661 TLLQDEIAEDYGEELGCMKLSSEIPLTLHSVSELVRLCLRRSDVQEESEGGSDTDNDKSAAP 720
 DB 661 TLLQDEIAEDYGEELGCMKLSSEIPLTLHSVSELVRLCLRRSDVQEESEGGSDTDNDKSAAP 720
 QY 721 DNEVQDFLEKLTSEPFELTSEKQLIILTALCHRLIMTYSVQDHMETRQMSAELWKER 780
 DB 721 DNEVQDFLEKLTSEPFELTSEKQLIILTALCHRLIMTYSVQDHMETRQMSAELWKER 780
 QY 781 LAVLKEENDKRAEKQKREKAKENKENGKVENGLGKTDKRIIVKFPQVDTEADMTS 840
 DB 781 LAVLKEENDKRAEKQKREKAKENKENGKVENGLGKTDKRIIVKFPQVDTEADMTS 840
 QY 841 AVKSRRLAIQAKKEREIOEREMVKLEBRQAEERIIRKHAIAAEKAFQEGIAKALVMRR 900
 DB 841 AVKSRRLAIQAKKEREIOEREMVKLEBRQAEERIIRKHAIAAEKAFQEGIAKALVMRR 900
 QY 901 TPIGTDRNHNRYLWFSDEVVPLFTIEKGWVHDSIDYRFNHHCKDHTVSGDEDYCPRSKXAN 960
 DB 901 TPIGTDRNHNRYLWFSDEVVPLFTIEKGWVHDSIDYRFNHHCKDHTVSGDEDYCPRSKXAN 960
 QY 961 LGKASMTQGTATEVAVETTPKQGNLFLCDSQKELDELNCLHPQIGRESQLKER 1020
 DB 961 LGKASMTQGTATEVAVETTPKQGNLFLCDSQKELDELNCLHPQIGRESQLKER 1020
 QY 1021 LEKRYODIIHSIHLARKPNLGLKSCDGNQELLNFLRSDLIEVATRLQKGGLYVETSEF 1080
 DB 1021 LEKRYODIIHSIHLARKPNLGLKSCDGNQELLNFLRSDLIEVATRLQKGGLYVETSEF 1080
 QY 1081 EARVISLEKLDGFCVIALQASVIKFLQGFMAPKQKRLQLQSDSASKEEVEDEBKQV 1140
 DB 1081 EARVISLEKLDGFCVIALQASVIKFLQGFMAPKQKRLQLQSDSASKEEVEDEBKQV 1140
 QY 1141 BEAKVASALEKWKTAIREAQTFSRMHVLLGMLDACIKWDSANENARCKVCPKKGDDKLI 1200
 DB 1141 BEAKVASALEKWKTAIREAQTFSRMHVLLGMLDACIKWDSANENARCKVCPKKGDDKLI 1200
 QY 1201 LCDECNKAFHLFLCLRPALYVEPDGWSQCPACQATARNRSGRNYTESASDSDESD 1260
 DB 1201 LCDECNKAFHLFLCLRPALYVEPDGWSQCPACQATARNRSGRNYTESASDSDESD 1260
 QY 1261 EEEEEEEEEEEDEYEVAGLRLRPKTIIRGKHSVIPPAARSGRRPGKXPHSTRSQKAP 1320
 DB 1261 EEEEEEEEEEEDEYEVAGLRLRPKTIIRGKHSVIPPAARSGRRPGKXPHSTRSQKAP 1320
 QY 1321 PVDDAEVDELVLQTKRSSRROSLELQKCEEILHKIVKRYFSPFPPTPRDEAEYDVI 1380
 DB 1321 PVDDAEVDELVLQTKRSSRROSLELQKCEEILHKIVKRYFSPFPPTPRDEAEYDVI 1380
 QY 1381 THPMDFTQVONKSCSGSYRSVQEFUTDMKQFTNAEVYNCRSHVLSCHWVKEQCLVILL 1440

Db 1381 THPMDFQTQVONKCSGYSVQBFLLDMKQVFNABEYVNCRSHVLSVCMVKTEQCLVALL 1440
QY 1441 HKHLPCHPVVRRKKKFPDRLAEDGDSPEAVGQSRDDEDRSRE 1485
Db 1441 HKHLPCHPVVRRKKKFPDRLAEDGDSPEAVGQSR--GRRQKK 1483

RESULT 4
ID ADF69140
ADP69140 standard; protein; 1483 AA.
XX ADF69140;
AC ADF69140;
XX 12-FEB-2004 (first entry)
DT Human MP53 protein sequence SEQ ID NO:110.
XX p53 pathway modulating agent; MP53; p53 modulator; cytostatic;
KW Gene therapy; cancer; human.
XX Homo sapiens.
XX WO2003083047-A2.
XX 09-OCT-2003.
XX 28-FEB-2003; 2003WO-US006025.
XX 01-MAR-2002; 2002US-0361196P.
XX (EXEL-) EXELIXIS INC. -
XX Belvin M, Francis-Lang H, Friedman L, Plowman GD, Heuer TS, Li D;
PI Funke RP;
XX WPI; 2003-812540/76.
DR N-PSDB; ADF69196.
XX

Identifying a candidate p53 pathway modulating agent for treating e.g.,
PT cancer by contacting an assay system comprising a MP53 polypeptide or
PT nucleic acid with a test agent and detecting a test agent-biased
PT activity.
XX Example; SEQ ID NO 110; 406pp; English.
XX The present invention describes a method for identifying a candidate p53
CC pathway modulating agent, which comprises: (a) providing an assay system
CC comprising a MP53 (modulator of p53) polypeptide or nucleic acid or its
CC fragment or derivative; (b) contacting the assay system with a test agent
CC under conditions where the system provides a reference activity except in
CC the presence of the test agent; and (c) detecting a test agent-biased
CC activity, where a difference between the test agent-biased activity and
CC the reference activity identifies the test agent as a candidate p53
CC pathway modulating agent. Also described: (1) modulating the p53 pathway
CC of a cell; (2) modulating the p53 pathway in a mammalian cell; and (3)
CC diagnosing a disease in a patient. MP53 has cytostatic activity, and can
CC be used in gene therapy. The method is useful for identifying a candidate
CC p53 pathway modulating agent for preparing a composition for diagnosing
CC or treating e.g., cancer. The present sequence represents a human MP53
CC protein, which is used in the exemplification of the present invention.
XX Sequence 1483 AA;
Query Match 96.2%; Score 7668; DB 7; Length 1483;
Best Local Similarity 99.2%; Pred. No. 0;
Matches 1473; Conservative 2; Mismatches 8; Indels 2; Gaps 1;
QY 1 MAPLLGRKFPVLVNPPLGPEEPFTTIPHTQEAFTREEYEARELYSERIWTCKSTGSSOL 60
Db 1 MAPLLGRKFPVLVNPPLGPEEPFTTIPHTQEAFTREEYEARELYSERIWTCKSTGSSOL 60
QY 61 THKEAWEEQVAEILLKEEFPAPWEKLVLEWMVHNTASLEKLVDTAWLEIMTKYAVGEEC 120

Db 61 THKEAWEEQVAEILLKEEFPAPWEKLVLEWMVHNTASLEKLVDTAWLEIMTKYAVGEEC 120
QY 121 DFEVGEKEMLVKVIKIHLEKVDREATEKKGDCGCDSPSSDKENSSQIAQDHQKKEVTV 180
Db 121 DFEVGEKEMLVKVIKIHLEKVDREATEKKGDCGCDSPSSDKENSSQIAQDHQKKEVTV 180
QY 181 KEDEGRRESINDRARRSPKLPSTLKKGERKWAPPKFLPHKYDVVKLQNEDKIISNPADS 240
Db 181 KEDEGRRESINDRARRSPKLPSTLKKGERKWAPPKFLPHKYDVVKLQNEDKIISNPADS 240
QY 241 LIRTERPPNKEIVRYFIRHNALRAGTGENAPWVVEDELVKYKLSLPSKFSDFLLDPKYMT 300
Db 241 LIRTERPPNKEIVRYFIRHNALRAGTGENAPWVVEDELVKYKLSLPSKFSDFLLDPKYMT 300
QY 301 LNPSTKRKNTGSPDRKPSKSKTDNSSLSPLNPKLWCHVHLKSLSGSPKLVKNSKNSK 360
Db 301 LNPSTKRKNTGSPDRKPSKSKTDNSSLSPLNPKLWCHVHLKSLSGSPKLVKNSKNSK 360
QY 361 SPEEHLEEMKMWSPNKLHTNFHIPKGPAPKPKGKHSKDKPLKAGRSKGILNGQKSTGN 420
Db 361 SPEEHLEEMKMWSPNKLHTNFHIPKGPAPKPKGKHSKDKPLKAGRSKGILNGQKSTGN 420
QY 421 SKSPKKGLKTPKTKMQLTLDMAKGTQKMTAPRNSGGTPTRTSSKPHGHLPPAALHLIA 480
Db 421 SKSPKKGLKTPKTKMQLTLDMAKGTQKMTAPRNSGGTPTRTSSKPHGHLPPAALHLIA 480
QY 481 YYKENKDRKRSALSCVISTARLLSSDRARLPEELRSLVQKRYELLEHKRWASMSSE 540
Db 481 YYKENKDRKRSALSCVISTARLLSSDRARLPEELRSLVQKRYELLEHKRWASMSSE 540
QY 541 EQRKEYLKKREELKKLKEKAKERREKEMLEKQKRYEQELTGKLPAPRLVDTPE 600
Db 541 EQRKEYLKKREELKKLKEKAKERREKEMLEKQKRYEQELTGKLPAPRLVDTPE 600
QY 601 GLPNTLFGDVAMVVEFLSCYSGLLLPDAQYPTAVSLMBEALADKGGFFLYLNRVLVILQ 660
Db 601 GLPNTLFGDVAMVVEFLSCYSGLLLPDAQYPTAVSLMBEALADKGGFFLYLNRVLVILQ 660
QY 661 TLQDEIAEDYGEKMKLSEIPLTLHVSSELVRLCLRRSDVOESEGSDTDNDKDSAAFE 720
Db 661 TLQDEIAEDYGEKMKLSEIPLTLHVSSELVRLCLRRSDVOESEGSDTDNDKDSAAFE 720
QY 721 DNEVQDEFLEKLETSEFFELTSEEEKLIQILTALCHRLMTYSVQDHMETRQMSAELWKER 780
Db 721 DNEVQDEFLEKLETSEFFELTSEEEKLIQILTALCHRLMTYSVQDHMETRQMSAELWKER 780
QY 781 LAVLKEENDKGAABKQKKEAKKENGKGVENGKGTDRKRIIVKFEPOVDTAEADMTIS 840
Db 781 LAVLKEENDKGAABKQKKEAKKENGKGVENGKGTDRKRIIVKFEPOVDTAEADMTIS 840
QY 841 AVKSRRLALIAQAKKEREIQEREMKVKLRQAEEERIIRKHAAAEKAFQEGIAKAKLVMR 900
Db 841 AVKSRRLALIAQAKKEREIQEREMKVKLRQAEEERIIRKHAAAEKAFQEGIAKAKLVMR 900
QY 901 TPIGTRDNHNRVYLFSDVPGLFIEKGWVHDSIDYRFNHHCKDHTVSGDEDYCPRSSKAN 960
Db 901 TPIGTRDNHNRVYLFSDVPGLFIEKGWVHDSIDYRFNHHCKDHTVSGDEDYCPRSSKAN 960
QY 961 LGKNASMTQHGTA TEAVAVETTPKQGNLWFLCDSQKELDELLNCLHPQIGRESOLKER 1020
Db 961 LGKNASMTQHGTA TEAVAVETTPKQGNLWFLCDSQKELDELLNCLHPQIGRESOLKER 1020
QY 1021 LEKRYQDIHSHIHLARKPNLGLKSCDGNQELANFLRSDLIEVATRLQKGLGVVETSEF 1080
Db 1021 LEKRYQDIHSHIHLARKPNLGLKSCDGNQELANFLRSDLIEVATRLQKGLGVVETSEF 1080
QY 1081 EARVLSLEKLDKDFGECVIALQASVKKFLQGFMAPKQKRRKQSDSDSAKTEVDEBKQKV 1140
Db 1081 EARVLSLEKLDKDFGECVIALQASVKKFLQGFMAPKQKRRKQSDSDSAKTEVDEBKQKV 1140
QY 1141 BEAKVASALEKWKTAIRBAQTFSRMHVLLGMLDACIKWDMSEANARCKVCPKKGDDKUI 1200

Db 1141 BEAKVASALEKWKTAIREAQTFSRMHVLLGLMDACIKWMSAENARCKVCRKKGDDDKLI 1200
QY 1201 LCDECNKAFHLECLRPALYEPDGEWQCPACQPATARNRSGRNTYESASEDSDESD 1260
Db 1201 LCDECNKAFHLECLRPALYEPDGEWQCPACQPATARNRSGRNTYESASEDSDESD 1260
QY 1261 BEEBEEEEEEDEVEAGLRPRKTIIRGKHSVIPPAAARSGRRPGKPHSTRSQKPAP 1320
Db 1261 BEEBEEEEEEDEVEAGLRPRKTIIRGKHSVIPPAAARSGRRPGKPHSTRSQKPAP 1320
QY 1321 PVDVAEDELVLQTRSSRRQSLQKCEIILHKIKVKYRFSWPPFVTRDEAEYDVI 1380
Db 1321 PVDVAEDELVLQTRSSRRQSLQKCEIILHKIKVKYRFSWPPFVTRDEAEYDVI 1380
QY 1381 THPMDQTVQNKCSGYSRVOEFLTDMKQVPTNAEVNCRGSHVLSQWVTEQCLVLVL 1440
Db 1381 THPMDQTVQNKCSGYSRVOEFLTDMKQVPTNAEVNCRGSHVLSQWVTEQCLVLVL 1440
QY 1441 HKHLPCHPVVRKRRKFPDLRLAEDSGDSEPEAVGOSRDEDRSRE 1485
Db 1441 HKHLPCHPVVRKRRKFPDLRLAEDSGDSEPEAVGOSR--GRQKK 1483

RESULT 5

ID ABR41364 standard; protein; 1052 AA.

AC ABR41364;

XX 02-JUN-2003 (first entry)

XX Human DITHP transcription factor.

KW Human; dithp; diagnostic and therapeutic polynucleotide; diagnosis;
KW cancer; cell proliferative disorder; autoimmune disorder;
KW inflammatory disorder; infection; hormonal disorder; metabolic disorder;
KW neurological disorder; gastrointestinal disorder; transport disorder;
KW connective tissue disorder; drug screening; proteome analysis;
KW gene therapy; anticense therapy; genotyping; transgenic animal; knock in;
KW disease model; toxicological testing; transcript imaging;
KW transcription factor.

XX Homo sapiens.

PN WO200297031-A2.

XX 05-DEC-2002.

XX 27-MAR-2002; 2002WO-US010056.

XX 28-MAR-2001; 2001US-0279619P.

PR 29-MAR-2001; 2001US-0280067P.

PR 29-MAR-2001; 2001US-0280068P.

PR 16-MAY-2001; 2001US-0291280P.

PR 17-MAY-2001; 2001US-0291829P.

PR 17-MAY-2001; 2001US-0291849P.

PR 19-JUN-2001; 2001US-0299428P.

PR 20-JUN-2001; 2001US-0299776P.

XX (INCY-) INCYTE GENOMICS INC.

XX Daffo A, Jones AL, Tran AB, Dahl CR, Gietzen D, Chinn J;

PI Dufour GE, Hillman JL, Yu JY, Tuason O, Yap PE, Amshey SR;

PI Daughtery SC, Dam TC, Liu TF, Nguyen DA, Kleefeld Y, Gerstin EH;

PI Peralta CH, David MH, Lewis SA, Chen AJ, Panzer SR, Harris B;

PI Flores V, Marwaha R, Lo A, Lan RY, Urashka ME;

XX WPI; 2003-129518/12.

PT diagnostic and therapeutic polynucleotide, and to induce antibodies.
XX Claim 27; SEQ ID NO 899; 591pp; English.

CC The invention relates to novel human diagnostic and therapeutic
CC polynucleotides designated dithp (ACC46080-ACC46749) and to their encoded
CC proteins (DITHP; ABR41136-ABR41812). The invention also relates to
CC polynucleotide sequences at least 90% identical to the dithp cDNA
CC sequences of the invention; recombinant vectors, host cells and
CC transgenic organisms comprising a dithp nucleic acid sequence; the
CC recombinant production of DITHP proteins; antibodies specific for DITHP
CC proteins; microarrays comprising dithp nucleic acid sequences; methods of
CC detecting dithp nucleotide and protein sequences; methods of screening
CC for compounds which specifically bind a DITHP protein; and methods of
CC assessing the toxicity of test compounds using a dithp hybridisation
CC probe. Dithp nucleic acid sequences and DITHP proteins may be used in the
CC diagnosis of a wide variety of conditions including cancer and other cell
CC proliferative disorders; autoimmune or inflammatory disorders; bacterial,
CC viral, fungal or parasitic infections; hormonal disorders; metabolic
CC disorders; neurological disorders; gastrointestinal disorders; transport
CC disorders; and connective tissue disorders. They may also be used to
CC screen for modulators of protein activity or gene expression. DITHP
CC proteins can additionally be used in analysis of the proteome of a tissue
CC or cell type and to induce antibodies. The dithp nucleic acids are
CC additionally useful in somatic or germline gene therapy of the disorders
CC mentioned above, as a source of antisense sequences, as a source of
CC probes and primers, in genotyping and identification of individuals, in
CC the generation of transgenic animal models of human disease or knock in
CC humanised animals, in toxicological testing, and in transcript imaging.
CC The present sequence represents a DITHP protein which has transcription
CC factor activity. Note: The sequence data for this patent did not form
CC part of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX Sequence 1052 AA;

Query Match 66.8%; Score 5319; DB 6; Length 1052;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1034; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 140 LEKVDDEATEKSDGACDPSDDKENSQIAQDHQKKEVVKEDGRRRESINDRARRSPR 199

Db 7 LEKVDEATERKSDGACDPSDDKENSQIAQDHQKKEVVKEDGRRRESINDRARRSPR 66

QY 200 KLPTSLKKGERKWAPPKELPHKYDVKLQNEDKIISNVPADSLIRTERPNKEIVRYFIRH 259

Db 67 KLPTSLKKGERKWAPPKELPHKYDVKLQNEDKIISNVPADSLIRTERPNKEIVRYFIRH 126

QY 260 NALRAGTGENAPWVVEDELVKYSLPSKFSDFLLDPYKMTLNPSTKKNKGSPDRKPSK 319

Db 127 NALRAGTGENAPWVVEDELVKYSLPSKFSDFLLDPYKMTLNPSTKKNKGSPDRKPSK 186

QY 320 KSKTDNSSLSPNPKLWCHVHLKKSLSGSPKLVKNKSNKSPPEHLEMMKMSPNKLH 379

Db 187 KSKTDNSSLSPNPKLWCHVHLKKSLSGSPKLVKNKSNKSPPEHLEMMKMSPNKLH 246

QY 380 TNFHIPKKGPPAKPGKHSDFPLKAKGRSGKILNGQKSTGNSKSPKGLKTPKTQKQMT 439

Db 247 TNFHIPKKGPPAKPGKHSDFPLKAKGRSGKILNGQKSTGNSKSPKGLKTPKTQKQMT 306

QY 440 LLDWAKGTQKMTAPRNSGGTPTSSKPHKHLPPAALHLIAYYKKNKREDKRSALSCVI 499

Db 307 LLDWAKGTQKMTAPRNSGGTPTSSKPHKHLPPAALHLIAYYKKNKREDKRSALSCVI 366

QY 500 SKTARLLSSDRARLPEELRSLVQKRYELLBHKKRWASMSSEQRKEYLKKKREELKKKLLK 559

Db 367 SKTARLLSSDRARLPEELRSLVQKRYELLBHKKRWASMSSEQRKEYLKKKREELKKKLLK 426

QY 560 EKAKERREKEMLERLEKQKRYEDQBLTGKNLPAFLVDTPEGLPNTLFGDVAMVVEFLSC 619

Db 427 EKAKERREKEMLERLEKQKRYEDQBLTGKNLPAFLVDTPEGLPNTLFGDVAMVVEFLSC 486

QY 620 YSGLLLPDAQYPIITAVSLMEALSADKGGFLYNRVLLVILLQTLQDEIAEDYGELGMKLK 679

Db 487 YSGLLPDAQYPTAVSLMEALSADGGFLYNRLVILLQTLLODEIAEDYGEIOMKLS 546
Qy 680 EIPLTLSVSELVRLCLRSVQEESESGSDTDNKSAAAFEDNEVQDEPLEKLTSEFFE 739
Db 547 EIPLTLSVSELVRLCLRSVQEESESGSDTDNKSAAAFEDNEVQDEPLEKLTSEFFE 606
Qy 740 LTSEKQLQTLTALCHRLMTYSVQDHMETROQMSAELWKRERLAVLKEENDKRAEKQK 799
Db 607 LTSEKQLQTLTALCHRLMTYSVQDHMETROQMSAELWKRERLAVLKEENDKRAEKQK 666
Qy 800 ENEAKNKENGKVENGLKTRKRIKVPPEQVDTEADMIKAVKSRLLAIQAKKERETQ 859
Db 667 ENEAKNKENGKVENGLKTRKRIKVPPEQVDTEADMIKAVKSRLLAIQAKKERETQ 726
Qy 860 EREMKVLRQAEERIRKHAAAEKAFQEGIAKAKLVNRRTPIGTDNRHNNRYWLFSD 919
Db 727 EREMKVLRQAEERIRKHAAAEKAFQEGIAKAKLVNRRTPIGTDNRHNNRYWLFSD 786
Qy 920 PGLFTEKGWVHDSIDYRFNHHCKDHTVSGDEDYCPRSKCANLGKSNMNTQHGTA 979
Db 787 PGLFTEKGWVHDSIDYRFNHHCKDHTVSGDEDYCPRSKCANLGKSNMNTQHGTA 846
Qy 980 ETTTPKQGNLWFLCDSQKELDELLNCLHPQIGRESQKXERLEKRYQDIHSHLARK 1039
Db 847 ETTTPKQGNLWFLCDSQKELDELLNCLHPQIGRESQKXERLEKRYQDIHSHLARK 906
Qy 1040 LGLKSCDGNQELLNFLRSDLIEVATRLQKGLGYVEETSEFEARVISLEKLD 1099
Db 907 LGLKSCDGNQELLNFLRSDLIEVATRLQKGLGYVEETSEFEARVISLEKLD 966
Qy 1100 LQASVKKFQLOGFMAPKQRRKLQSEDSAKTEEVDEEKQWBEAKVASALEKWKTA 1159
Db 967 LQASVKKFQLOGFMAPKQRRKLQSEDSAKTEEVDEEKQWBEAKVASALEKWKTA 1026
Qy 1160 QTFSRMHLGLMDA 1174
Db 1027 QTFSRMHLGLMDA 1041

RESULT 6
AAM14838
ID AAM14838 standard; protein; 572 AA.
XX
AC AAM14838;
XX
DT 12-OCT-2001 (first entry)
XX
DE Peptide #1272 encoded by probe for measuring cervical gene expression.
XX
KW Probe; human; microarray; gene expression; cervical epithelial cell;
KW cervical cancer.
XX
OS Homo sapiens.
XX
PN WO200157278-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US000670.
XX
PR 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX

DR WPI; 2001-488901/53.
XX Human genome-derived single exon nucleic acid probes useful for analyzing
PT gene expression in human cervical epithelial cells.
XX
PS Claim 27; SEQ ID NO 19664; 487pp; English.
XX
CC The present invention relates to human single exon nucleic acid probes
CC (SENP: see AAI10068-AAI28459). The present sequence is a peptide encoded
CC by one such probe. The SENPs are derived from human Hela cells. The SENPs
CC can be used to produce a single exon microarray, which can be used for
CC measuring human gene expression in a sample derived from human cervical
CC epithelial cells. By measuring gene expression, the probes are therefore
CC useful in grading and/or staging of diseases of the cervix, notably
CC cervical cancer. Note: The sequence data for this patent did not form
CC part of the printed specification, but was obtained in electronic format
CC directly from WIPO at fcp.wipo.int/pub/published_pct_sequences
XX Sequence 572 AA;
SQ

Query Match 36.3%; Score 2889; DB 4; Length 572;
Best Local Similarity 99.5%; Pred. No. 2.5e-185;
Matches 567; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 298 YMTLPSTKRNKGTGSPDRKPSKSKTDNSSLSPNPKLWCHVHLKSLSGSPLKVNK 357
Db 1 YMTLPSTKRNKGTGSPDRKPSKSKTDNSSLSPNPKLWCHVHLKSLSGSPLKVNK 60
Qy 358 NSKSPPEEHEEMKQMSPNKLTNPHIPKGGPAKPGKSDPKLAKAGRSKGLNGQKS 417
Db 61 NSKSPPEEHEEMKQMSPNKLTNPHIPKGGPAKPGKSDPKLAKAGRSKGLNGQKS 120
Qy 418 TGNSSPKKGLTKPTKMKQMTLLDMAKGTQKWTAPRNSGCTPRTSSKPHKLPALH 477
Db 121 TGNSSPKKGLTKPTKMKQMTLLDMAKGTQKWTAPRNSGCTPRTSSKPHKLPALH 180
Qy 478 LIAYYKENKDRKRSALSCVISKTARLLSSDRARLPEELRSVQRYELLEHKKRWAS 537
Db 181 LIAYYKENKDRKRSALSCVISKTARLLSSDRARLPEELRSVQRYELLEHKKRWAS 240
Qy 538 MSEORKEYLKKREELKKLKEKAKERREKEMLEKQKEYEDELTKNLPAPRLVD 597
Db 241 MSEORKEYLKKREELKKLKEKAKERREKEMLEKQKEYEDELTKNLPAPRLVD 300
Qy 598 TPEGLPNTLFGDVAMVVEFLSCYSGLLLPDAQYPTAVSLMEALSADKGGFLYLNRLVI 657
Db 301 TPEGLPNTLFGDVAMVVEFLSCYSGLLLPDAQYPTAVSLMEALSADKGGFLYLNRLVI 360
Qy 658 LIQTLQDEIAEDYGEIOMKLSIEIPTLHVSSELVRLCLRRSDVQEESESGSDTDNKS 717
Db 361 LIQTLQDEIAEDYGEIOMKLSIEIPTLHVSSELVRLCLRRSDVQEESESGSDTDNKS 420
Qy 718 APEDNEVQDEPLEKLTSEFFELTSEKLOITLTALCHRLMTYSVQDHMETROQMSAEL 777
Db 421 APEDNEVQDEPLEKLTSEFFELTSEKLOITLTALCHRLMTYSVQDHMETROQMSAEL 480
Qy 778 KERLAVLKEENDKRAEKQKRAEKQKRAEKQKRAEKQKRAEKQKRAEKQKRAEKQ 837
Db 481 KERLAVLKEENDKRAEKQKRAEKQKRAEKQKRAEKQKRAEKQKRAEKQKRAEKQ 480
Qy 838 MISAVKSRLLAIQAKKERETQEREMKVKL 867
Db 541 MISAVKSRLLAIQAKKERETQEREMKVKI 570

RESULT 7
ABB33805
ID ABB33805 standard; peptide; 572 AA.
XX
AC ABB33805;
XX
DT 04-FEB-2002 (first entry)
XX

DE Peptide #1311 encoded by human foetal liver single exon probe.
 XX Human; foetal liver; gene expression; single exon nucleic acid probe.
 KW Homo sapiens.
 OS WO200157277-A2.
 PN 09-AUG-2001.
 PD 30-JAN-2001; 2001WO-US000669.
 XX 04-FEB-2000; 2000US-0180312P.
 XX 26-MAY-2000; 2000US-0207456P.
 PR 30-JUN-2000; 2000US-00608408.
 PR 03-AUG-2000; 2000US-00632366.
 PR 21-SEP-2000; 2000US-0234687P.
 PR 27-SEP-2000; 2000US-0236359P.
 PR 04-OCT-2000; 2000GB-00024263.
 XX (MOLE-) MOLECULAR DYNAMICS INC.
 PA Penn SG, Hanzel DK, Chen W, Rank DR;
 XX WPI; 2001-483447/52.
 DR Human genome-derived single exon nucleic acid probes useful for analyzing
 XX gene expression in human fetal liver.
 PT Claim 27; SEQ ID NO 26440; 639pp + Sequence Listing; English.
 PS The invention relates to a single exon nucleic acid probe for measuring
 XX human gene expression in a sample derived from human foetal liver. The
 CC single exon nucleic acid probes may be used for predicting, measuring and
 CC displaying gene expression in samples derived from human fetal liver. The
 CC present sequence is a peptide encoded by a single exon nucleic acid probe
 CC of the invention. Note: The sequence data for this patent did not form
 CC part of the printed specification, but was obtained in electronic format
 CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX Sequence 572 AA;
 SQ
 Query Match 36.3%; Score 2889; DB 4; Length 572;
 Best Local Similarity 99.5%; Pred. No. 2.5e-185;
 Matches 567; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 QY 298 YMTLNPSTKRKNTGSPDRKPKSKKTDNSSLSPLNPKLWCHVHLKXLSGSPKVKNSK 357
 DB 1 YMTLNPSTKRKNTGSPDRKPKSKKTDNSSLSPLNPKLWCHVHLKXLSGSPKVKNSK 60
 QY 358 NSKSPPEHLEEMWKMSPNKLHNFHIPKGGPPAKKPGKHSKDKPLKAKGRSGILNGOKS 417
 DB 61 NSKSPPEHLEEMWKMSPNKLHNFHIPKGGPPAKKPGKHSKDKPLKAKGRSGILNGOKS 120
 QY 418 TGNKSPKGLTKPTKTKQMTLLDMAGTQKWTAPRNSGTPRTSKPKHLPAAIH 477
 DB 121 TGNKSPKGLTKPTKTKQMTLLDMAGTQKWTAPRNSGTPRTSKPKHLPAAIH 180
 QY 478 LIAYYKENKDRDKSALSVCVISTARLLSSDRARLPEELRSIVQKYLELLEHKRWAS 537
 DB 181 LIAYYKENKDRDKSALSVCVISTARLLSSDRARLPEELRSIVQKYLELLEHKRWAS 240
 QY 538 MSEQRKYLKKRBEELKKLKEKAKERREKEMLERLEKQRYEDQELTGKNLPAPRLVD 597
 DB 241 MSEQRKYLKKRBEELKKLKEKAKERREKEMLERLEKQRYEDQELTGKNLPAPRLVD 300
 QY 598 TPEGIPNTLFGDVAMVVEFLSCYGLLPDAQYPTAVSLMEALSADKGGFLYNRLVI 657
 DB 301 TPEGIPNTLFGDVAMVVEFLSCYGLLPDAQYPTAVSLMEALSADKGGFLYNRLVI 360
 QY 658 LIQTLLOEIAEDYELGKMLSEIPLTLHSVSELVRLCLRRSDVQESGSDTDNDKSA 717
 DB 361 LIQTLLOEIAEDYELGKMLSEIPLTLHSVSELVRLCLRRSDVQESGSDTDNDKSA 420

QY 718 APEDNEVODEFLEKLETSEFELTSEEKLOILTALCHRLMTYSVQDHMETRQOMSABLW 777
 DB 421 APEDNEVODEFLEKLETSEFELTSEEKLOILTALCHRLMTYSVQDHMETRQOMSABLW 480
 QY 778 KERLAVLKEENDKKRAEKQKREKEMAKNKENGKVENGLGKTDRKKRIVKFFPQVDTEAED 837
 DB 481 KERLAVLKEENDKKRAEKQKREKEMAKNKENGKVENGLGKTDRKKRIVKFFPQVDTEAED 540
 QY 838 MISAVKSRLLAIQAKKEREIOEREMKVKL 867
 DB 541 MISAVKSRLLAIQAKKEREIOEREMKGI 570
 RESULT 8
 AAM27265
 ID AAM27265 standard; protein; 572 AA.
 XX AC AAM27265;
 XX DT 17-OCT-2001 (first entry)
 XX DE Peptide #1302 encoded by probe for measuring placental gene expression.
 XX KW Probe; microarray; human; placenta; antenatal diagnosis;
 XX KW Genetic disorder.
 XX OS Homo sapiens.
 XX PN WO200157272-A2.
 XX PD 09-AUG-2001.
 XX PF 30-JAN-2001; 2001WO-US000663.
 XX PR 04-FEB-2000; 2000US-0180312P.
 XX PR 26-MAY-2000; 2000US-0207456P.
 XX PR 30-JUN-2000; 2000US-00608408.
 XX PR 03-AUG-2000; 2000US-00632366.
 XX PR 21-SEP-2000; 2000US-0234687P.
 XX PR 27-SEP-2000; 2000US-0236359P.
 XX PR 04-OCT-2000; 2000GB-00024263.
 XX (MOLE-) MOLECULAR DYNAMICS INC.
 PA Penn SG, Hanzel DK, Chen W, Rank DR;
 XX WPI; 2001-488897/53.
 XX Human genome-derived single exon nucleic acid probes useful for analyzing
 XX gene expression in human placenta.
 XX PS Claim 27; SEQ ID NO 27534; 654pp; English.
 XX CC The present invention relates to single exon nucleic acid probes (SENPs;
 CC see AA13135-AA157546). The present sequence is a peptide encoded by one
 CC such probe. The probes are useful for producing a microarray for
 CC predicting, measuring and displaying gene expression in samples derived
 CC from human placenta. The probes are useful for antenatal diagnosis of
 CC human genetic disorders
 XX SQ Sequence 572 AA;
 Query Match 36.3%; Score 2889; DB 4; Length 572;
 Best Local Similarity 99.5%; Pred. No. 2.5e-185;
 Matches 567; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 QY 298 YMTLNPSTKRKNTGSPDRKPKSKKTDNSSLSPLNPKLWCHVHLKXLSGSPKVKNSK 357
 DB 1 YMTLNPSTKRKNTGSPDRKPKSKKTDNSSLSPLNPKLWCHVHLKXLSGSPKVKNSK 60
 QY 358 NSKSPPEHLEEMWKMSPNKLHNFHIPKGGPPAKKPGKHSKDKPLKAKGRSGILNGOKS 417

Db 61 NSKSPHEHLEEMKMMSPNKLHTNFHPKGGPPAKPGKHSDKPLKAKGRSGKILNGQKS 120
Qy 418 TGNKSPKGLKTPKTKMQMTLLDMAKGTQKMTAPRNSGGTPTSSKPHKHLPPAALH 477
Db 121 TGNKSPKGLKTPKTKMQMTLLDMAKGTQKMTAPRNSGGTPTSSKPHKHLPPAALH 180
Qy 478 LIAYYKENDKREDKRSALSCVISTARLLSSDRARLPBELSLVQKRYELLEHKRWAS 537
Db 181 LIAYYKENDKREDKRSALSCVISTARLLSSDRARLPBELSLVQKRYELLEHKRWAS 240
Qy 538 MSEEORKEVYLKKREELKKLKEKAKERREKEMLEKQRYEDDELTKNLPAPRLVD 597
Db 241 MSEEORKEVYLKKREELKKLKEKAKERREKEMLEKQRYEDDELTKNLPAPRLVD 300
Qy 598 TPEGLPNTLFGDVAMVVEFLSCYSGLLLPDAQYPTAVSLMEALSADKGGFLYLNRLVI 657
Db 301 TPEGLPNTLFGDVAMVVEFLSCYSGLLLPDAQYPTAVSLMEALSADKGGFLYLNRLVI 360
Qy 658 LIQTLQDEIAEDYGEGLKLSSEIPIPLTHSVSELVRLCLRRSDVQESGSDTDDNKDSA 717
Db 361 LIQTLQDEIAEDYGEGLKLSSEIPIPLTHSVSELVRLCLRRSDVQESGSDTDDNKDSA 420
Qy 718 AFEDNEVDQEFLEKLETSFEFFELTSEEKLIQILTALCHRIILMTYSVODHMETROQMSAELW 777
Db 421 AFEDNEVDQEFLEKLETSFEFFELTSEEKLIQILTALCHRIILMTYSVODHMETROQMSAELW 480
Qy 778 KERLAVLKEENDKRAEKQKREKEMAKNKENGKVENGLGKTDRKKRIIVKFEQVDTAEAD 837
Db 481 KERLAVLKEENDKRAEKQKREKEMAKNKENGKVENGLGKTDRKKRIIVKFEQVDTAEAD 540
Qy 838 MISAVKSRRLAIQAKKEREIOEREMKVKL 867
Db 541 MISAVKSRRLAIQAKKEREIOEREMKVKL 570

RESULT 9
ABB28622
ID ABB28622 standard; peptide; 572 AA.
AC ABB28622;
XX
XX 01-FEB-2002 (first entry)
DT Peptide #1273 encoded by breast cell single exon nucleic acid probe.
XX Human; microarray; single exon probe; gene expression; breast; disease;
KW cancer.
XX
OS Homo sapiens.
XX
XX WO200157271-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US0000662.
XX
XX 04-FEB-2000; 2000US-0180312P.
XX 26-MAY-2000; 2000US-0207456P.
XX 30-JUN-2000; 2000US-00608408.
XX 03-AUG-2000; 2000US-00632366.
XX 21-SEP-2000; 2000US-0234687P.
XX 27-SEP-2000; 2000US-0236359P.
XX 04-OCT-2000; 2000GB-00024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-496933/54.
XX
PT New spatially-addressable set of single exon nucleic acid probes, useful
PT for measuring gene expression in sample derived from human breast,
PT comprises number of single exon nucleic acid probes.

XX
PS Claim 27; SEQ ID NO 11590; 327pp + Sequence Listing; English.
XX
CC The invention relates to a spatially-addressable set of single exon
CC nucleic acid probes for measuring gene expression in a sample derived
CC from human breast and BT 474 cells. The method involves contacting the
CC probes with a collection of detectably labelled nucleic acids derived
CC from mRNA of human breast, and then measuring the label bound to each
CC probe of the microarray. The probes are useful for verifying the
CC expression of regions of genomic DNA predicted to encode proteins. They
CC are useful for gene discovery, and for determining predisposition and/or
CC prognosing breast disease. Gene expression analysis is useful for
CC assessing the toxicity of chemical agents on cells. The microarray of
CC this invention presents a far greater diversity of probes for measuring
CC gene expression, with far less bias than expressed sequence tag
CC microarrays. The method is suitable for rapid production of functional
CC information from genomic sequence. The present sequence is a peptide
CC encoded by a single exon nucleic acid probe of the invention. Note: The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 572 AA;
Query Match 36.3%; Score 2889; DB 4; Length 572;
Best Local Similarity 99.5%; Pred. No. 2.5e-185;
Matches 567; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
Qy 298 YMTLNPSIKRNTGSPDRKPSKSKTDNSLSPLNPKLWCHVHLKSLSGSPLKYNK 357
Db 1 YMTLNPSIKRNTGSPDRKPSKSKTDNSLSPLNPKLWCHVHLKSLSGSPLKYNK 60
Qy 358 NSKSPKEHLEEMKMMSPNKLHTNFHPKGGPPAKPGKHSDKPLKAKGRSGKILNGQKS 417
Db 61 NSKSPKEHLEEMKMMSPNKLHTNFHPKGGPPAKPGKHSDKPLKAKGRSGKILNGQKS 120
Qy 418 TGNKSPKGLKTPKTKMQMTLLDMAKGTQKMTAPRNSGGTPTSSKPHKHLPPAALH 477
Db 121 TGNKSPKGLKTPKTKMQMTLLDMAKGTQKMTAPRNSGGTPTSSKPHKHLPPAALH 180
Qy 478 LIAYYKENDKREDKRSALSCVISTARLLSSDRARLPBELSLVQKRYELLEHKRWAS 537
Db 181 LIAYYKENDKREDKRSALSCVISTARLLSSDRARLPBELSLVQKRYELLEHKRWAS 240
Qy 538 MSEEORKEVYLKKREELKKLKEKAKERREKEMLEKQRYEDDELTKNLPAPRLVD 597
Db 241 MSEEORKEVYLKKREELKKLKEKAKERREKEMLEKQRYEDDELTKNLPAPRLVD 300
Qy 598 TPEGLPNTLFGDVAMVVEFLSCYSGLLLPDAQYPTAVSLMEALSADKGGFLYLNRLVI 657
Db 301 TPEGLPNTLFGDVAMVVEFLSCYSGLLLPDAQYPTAVSLMEALSADKGGFLYLNRLVI 360
Qy 658 LIQTLQDEIAEDYGEGLKLSSEIPIPLTHSVSELVRLCLRRSDVQESGSDTDDNKDSA 717
Db 361 LIQTLQDEIAEDYGEGLKLSSEIPIPLTHSVSELVRLCLRRSDVQESGSDTDDNKDSA 420
Qy 718 AFEDNEVDQEFLEKLETSFEFFELTSEEKLIQILTALCHRIILMTYSVODHMETROQMSAELW 777
Db 421 AFEDNEVDQEFLEKLETSFEFFELTSEEKLIQILTALCHRIILMTYSVODHMETROQMSAELW 480
Qy 778 KERLAVLKEENDKRAEKQKREKEMAKNKENGKVENGLGKTDRKKRIIVKFEQVDTAEAD 837
Db 481 KERLAVLKEENDKRAEKQKREKEMAKNKENGKVENGLGKTDRKKRIIVKFEQVDTAEAD 540
Qy 838 MISAVKSRRLAIQAKKEREIOEREMKVKL 867
Db 541 MISAVKSRRLAIQAKKEREIOEREMKVKL 570

RESULT 10
ABB19248
ID ABB19248 standard; protein; 572 AA.
XX


```
Query Match      36.3%; Score 2889; DB 4; Length 572;
Best Local Similarity 99.5%; Pred. No. 2.5e-185;
Matches 567; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 298 YMTLNSTKRNKTSPPDRKPSKSKTDNSLSPLNPKLWCHVHLKXSLSGSPKLVKNSK 357
    |||||
Db 1 YMTLNSTKRNKTSPPDRKPSKSKTDNSLSPLNPKLWCHVHLKXSLSGSPKLVKNSK 60

QY 358 NSKSPBEHLEEMKMMSPNKLHTNFHPPKGGPPAKPGKHSDKPLKAKGRSGKILNGQKS 417
    |||||
Db 61 NSKSPBEHLEEMKMMSPNKLHTNFHPPKGGPPAKPGKHSDKPLKAKGRSGKILNGQKS 120

QY 418 TGNKSPKKGKLTPTKTKMOMTLLDMAKGTQKTRAPRNSGGTPTRTSSKPHKLPAAALH 477
    |||||
Db 121 TGNKSPKKGKLTPTKTKMOMTLLDMAKGTQKTRAPRNSGGTPTRTSSKPHKLPAAALH 180

QY 478 LIAYYKENDREDKRSALSCVISTARTLLSSDRARLPEELSLVQRYELLEHKKRWAS 537
    |||||
Db 181 LIAYYKENDREDKRSALSCVISTARTLLSSDRARLPEELSLVQRYELLEHKKRWAS 240

QY 538 MSEEQRKEYLKKRBEELKKLKEKAKERREKEMLEKQRYEDQELTGKNLPAPRLVD 597
    |||||
Db 241 MSEEQRKEYLKKRBEELKKLKEKAKERREKEMLEKQRYEDQELTGKNLPAPRLVD 300

QY 598 TPEGLPNTLFGDVAMVVEFLSCYSGLLLPDAQYPTAVSLMEALSADKGGFLYLNRLVI 657
    |||||
Db 301 TPEGLPNTLFGDVAMVVEFLSCYSGLLLPDAQYPTAVSLMEALSADKGGFLYLNRLVI 360

QY 658 LIQTLQDEIAEDYGEIGMKLSEIPLTLHSVSELVRLCLRRSDVQESGSDTDDNKDSA 717
    |||||
Db 361 LIQTLQDEIAEDYGEIGMKLSEIPLTLHSVSELVRLCLRRSDVQESGSDTDDNKDSA 420

QY 718 AFEDNEVQDEFEKLETSEFFELTSEKLIQILTALCHRILMTYSVQDHMETROQMSAELW 777
    |||||
Db 421 AFEDNEVQDEFEKLETSEFFELTSEKLIQILTALCHRILMTYSVQDHMETROQMSAELW 480

QY 778 KERLAVLKEENDKKAERKAEKQKREKEMLEKQRYEDQELTGKTRKRIIVKPEQVDTEAD 837
    |||||
Db 481 KERLAVLKEENDKKAERKAEKQKREKEMLEKQRYEDQELTGKTRKRIIVKPEQVDTEAD 540

QY 838 MISAVKSRLLAIQAKEREIOEREMKVKL 867
    |||||
Db 541 MISAVKSRLLAIQAKEREIOEREMKVKL 570

RESULT 12
AAM54571
ID AAM54571 standard; protein; 572 AA.
XX AC AAM54571;
XX DT 05-NOV-2001 (first entry)
XX DE Human brain expressed single exon probe encoded protein SEQ ID NO: 26676.
XX KW Human; brain expressed exon; gene expression analysis; probe; microarray;
XX KW Alzheimer's disease; multiple sclerosis; schizophrenia; epilepsy; cancer.
XX OS Homo sapiens.
XX PN WO200157275-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US000667.
XX PR 04-FEB-2000; 2000US-0180312P.
XX PR 26-MAY-2000; 2000US-0207456P.
XX PR 30-JUN-2000; 2000US-00608408.
XX PR 03-AUG-2000; 2000US-00632366.
XX PR 21-SEP-2000; 2000US-0234687P.
XX PR 27-SEP-2000; 2000US-0236359P.
XX PR 04-OCT-2000; 2000GB-00024263.
```

```
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-483446/52.
XX Single exon nucleic acid probes for analyzing gene expression in human
XX brains.
XX Example 4; SEQ ID NO 26676; 650pp + Sequence Listing; English.
XX The present invention provides a number of single exon nucleic acid
XX probes which are derived from genomic sequences expressed in the human
XX brain. They can be used to measure gene expression in brain cell samples,
XX which may enable the diagnosis and improved treatment of nervous system
XX diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
XX epilepsy and cancers. The present sequence is a protein encoded by one of
XX the probes of the invention
XX
SQ Sequence 572 AA;
Query Match      36.3%; Score 2889; DB 4; Length 572;
Best Local Similarity 99.5%; Pred. No. 2.5e-185;
Matches 567; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 298 YMTLNSTKRNKTSPPDRKPSKSKTDNSLSPLNPKLWCHVHLKXSLSGSPKLVKNSK 357
    |||||
Db 1 YMTLNSTKRNKTSPPDRKPSKSKTDNSLSPLNPKLWCHVHLKXSLSGSPKLVKNSK 60

QY 358 NSKSPBEHLEEMKMMSPNKLHTNFHPPKGGPPAKPGKHSDKPLKAKGRSGKILNGQKS 417
    |||||
Db 61 NSKSPBEHLEEMKMMSPNKLHTNFHPPKGGPPAKPGKHSDKPLKAKGRSGKILNGQKS 120

QY 418 TGNKSPKKGKLTPTKTKMOMTLLDMAKGTQKTRAPRNSGGTPTRTSSKPHKLPAAALH 477
    |||||
Db 121 TGNKSPKKGKLTPTKTKMOMTLLDMAKGTQKTRAPRNSGGTPTRTSSKPHKLPAAALH 180

QY 478 LIAYYKENDREDKRSALSCVISTARTLLSSDRARLPEELSLVQRYELLEHKKRWAS 537
    |||||
Db 181 LIAYYKENDREDKRSALSCVISTARTLLSSDRARLPEELSLVQRYELLEHKKRWAS 240

QY 538 MSEEQRKEYLKKRBEELKKLKEKAKERREKEMLEKQRYEDQELTGKNLPAPRLVD 597
    |||||
Db 241 MSEEQRKEYLKKRBEELKKLKEKAKERREKEMLEKQRYEDQELTGKNLPAPRLVD 300

QY 598 TPEGLPNTLFGDVAMVVEFLSCYSGLLLPDAQYPTAVSLMEALSADKGGFLYLNRLVI 657
    |||||
Db 301 TPEGLPNTLFGDVAMVVEFLSCYSGLLLPDAQYPTAVSLMEALSADKGGFLYLNRLVI 360

QY 658 LIQTLQDEIAEDYGEIGMKLSEIPLTLHSVSELVRLCLRRSDVQESGSDTDDNKDSA 717
    |||||
Db 361 LIQTLQDEIAEDYGEIGMKLSEIPLTLHSVSELVRLCLRRSDVQESGSDTDDNKDSA 420

QY 718 AFEDNEVQDEFEKLETSEFFELTSEKLIQILTALCHRILMTYSVQDHMETROQMSAELW 777
    |||||
Db 421 AFEDNEVQDEFEKLETSEFFELTSEKLIQILTALCHRILMTYSVQDHMETROQMSAELW 480

QY 778 KERLAVLKEENDKKAERKAEKQKREKEMLEKQRYEDQELTGKTRKRIIVKPEQVDTEAD 837
    |||||
Db 481 KERLAVLKEENDKKAERKAEKQKREKEMLEKQRYEDQELTGKTRKRIIVKPEQVDTEAD 540

QY 838 MISAVKSRLLAIQAKEREIOEREMKVKL 867
    |||||
Db 541 MISAVKSRLLAIQAKEREIOEREMKVKL 570

RESULT 13
ABG48640
ID ABG48640 standard; peptide; 572 AA.
XX AC ABG48640;
XX
```

DT 25-FEB-2003 (first entry)

XX Human liver peptide, SEQ ID NO 27288.

DE Human; liver; cirrhosis; hyperlipoproteinaemia; hyperlipidaemia;

KW hypercholesterolaemia; coronary heart disease.

XX Homo sapiens.

OS

XX WO200157273-A2.

PN 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US000664.

XX 04-FEB-2000; 2000US-0180312P.

PR 26-MAY-2000; 2000US-0207456P.

PR 30-JUN-2000; 2000US-00608408.

PR 03-AUG-2000; 2000US-00632366.

PR 21-SEP-2000; 2000US-0234687P.

PR 27-SEP-2000; 2000US-0236359P.

PR 04-OCT-2000; 2000GB-00024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

PA Penn SG, Hanzel DK, Chen W, Rank DR;

PI WPI; 2001-488898/53.

XX Human genome-derived single exon nucleic acid probes useful for analyzing

PT gene expression in human adult liver.

XX Claim 27; SEQ ID NO 27288; 658pp; English.

XX The invention relates to a single exon nucleic acid probe (SENP) (I) for

CC measuring human gene expression in a sample derived from human adult

CC liver, comprising one of 13109 defined nucleotide sequences given in the

CC specification (or complements/ fragments). The probe hybridises at high

CC stringency to a nucleic acid molecule expressed in the human adult liver.

CC (I) may be used for predicting, measuring and displaying gene expression

CC in samples derived from human adult liver. The genes identified may be

CC involved in genetic liver diseases such as cirrhosis,

CC hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which is

CC associated with coronary heart disease. ABG47348-ABG59930 represent human

CC liver single exon encoded peptides of the invention. Note: the sequence

CC information for this patent does not appear in the printed specification

CC but was obtained in electronic format directly from WIPO at

CC ftp.wipo.int/pub/published_pct_sequences

XX Sequence 572 AA;

SQ

Query Match 36.3%; Score 2889; DB 4; Length 572;

Best Local Similarity 99.5%; Pred. No. 2.5e-185;

Matches 567; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 298 YMTLPSTKRNKTGSPDRKPSKSTDNSSLSPLNPKLWCHVHLKXSLSGSPKXVNSK 357

DB 1 YMTLPSTKRNKTGSPDRKPSKSTDNSSLSPLNPKLWCHVHLKXSLSGSPKXVNSK 60

QY 358 NSKSPPEHLEENMKWSPNKLHNFHPPKGGPPAKKPGKHSKPLKAGRSKGLNGOKS 417

DB 61 NSKSPPEHLEENMKWSPNKLHNFHPPKGGPPAKKPGKHSKPLKAGRSKGLNGOKS 120

QY 418 TGNKSPKPKGLTKPTKTKQMTLLDMAKGTQKWTAPRNSGTPRTSSKPHKLPALH 477

DB 121 TGNKSPKPKGLTKPTKTKQMTLLDMAKGTQKWTAPRNSGTPRTSSKPHKLPALH 180

QY 478 LIAYKENKORDKESKLSKVTLSKRTARLLSSDRARLPEELSLVOKRYELLEHKRWAS 537

DB 181 LIAYKENKORDKESKLSKVTLSKRTARLLSSDRARLPEELSLVOKRYELLEHKRWAS 240

QY 538 MSEQRKEYLKKREELKKLKEKAKERREKEMLERLEKQKRYEQDELTKGNLPAFLVD 597

DB 241 MSEQRKEYLKKREELKKLKEKAKERREKEMLERLEKQKRYEQDELTKGNLPAFLVD 300

QY 598 TPEGLPNTLFGDVAMVVEFLSCYSGLLLPDQYPTAVSLMEALSADKGGFLYLNRLVI 657

DB 301 TPEGLPNTLFGDVAMVVEFLSCYSGLLLPDQYPTAVSLMEALSADKGGFLYLNRLVI 360

QY 658 LLQTLQDEIAEDYGELGKMLSEIPLTLHVSSELVRLCLRRSDVQEESEGSDDTDNDKDSA 717

DB 361 LLQTLQDEIAEDYGELGKMLSEIPLTLHVSSELVRLCLRRSDVQEESEGSDDTDNDKDSA 420

QY 718 APEDNEVQDEFLEKLTSEPFELTSEEKLIQTLTALCHRLMTYSVQDHMETRQMSAELW 777

DB 421 APEDNEVQDEFLEKLTSEPFELTSEEKLIQTLTALCHRLMTYSVQDHMETRQMSAELW 480

QY 778 KERLAVLKEENDKKRAEKQKMEKKNKENGKVGNGKTKDRKKIYKFFSPQVDTAEAD 837

DB 481 KERLAVLKEENDKKRAEKQKMEKKNKENGKVGNGKTKDRKKIYKFFSPQVDTAEAD 540

QY 838 MISAVKSRLLAIQAKKEREIQEREMKVKL 867

DB 541 MISAVKSRLLAIQAKKEREIQEREMKVKI 570

RESULT 14

AA02563

ID AA02563 standard; protein; 572 AA.

XX AC AA02563;

XX 09-OCT-2001 (first entry)

DE Peptide #1245 encoded by probe for measuring breast gene expression.

XX Probe; human; breast disease; breast cancer; development disorder;

KW inflammatory disease; proliferative breast disease; non-carcinoma tumour.

XX Homo sapiens.

XX WO200157270-A2.

XX 09-AUG-2001.

XX 29-JAN-2001; 2001WO-US000661.

XX 04-FEB-2000; 2000US-0180312P.

PR 26-MAY-2000; 2000US-0207456P.

PR 30-JUN-2000; 2000US-00608408.

PR 03-AUG-2000; 2000US-00632366.

PR 21-SEP-2000; 2000US-0234687P.

PR 27-SEP-2000; 2000US-0236359P.

PR 04-OCT-2000; 2000GB-00024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

PA Penn SG, Hanzel DK, Chen W, Rank DR;

PI WPI; 2001-476286/51.

XX Novel single exon nucleic acid probe used to measuring gene expression in

PT a human breast.

XX Claim 27; SEQ ID NO 11303; 322pp; English.

XX The present invention relates to novel single exon nucleic acid probes

CC (see AA10010-AA11067). The present sequence is a peptide encoded by one

CC such probe. The probes are useful for measuring human gene expression in

CC a human breast sample, where the probe hybridises at high stringency to a

CC nucleic acid expressed in the human breast. The probes are useful for

CC predicting, diagnosing, grading, staging, monitoring and prognosing

CC diseases of the human breast, particularly those diseases with polygenic

CC aetiology. The diseases include: breast cancer, disorders of development,

CC inflammatory diseases of the breast, fibrocystic changes, proliferative

CC breast disease and non-carcinoma tumours. Note: The sequence data for

CC this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences

SQ Sequence 572 AA;

Query Match 36.3%; Score 2889; DB 4; Length 572;
Best Local Similarity 99.5%; Pred. No. 2.5e-185;
Matches 567; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 298 YMTLPSTKRKNTGSPDRKPSKSTDNSSPLNPKLWCHVHLKSLGSGPLKVNKSK 357
Db 1 YMTLPSTKRKNTGSPDRKPSKSTDNSSPLNPKLWCHVHLKSLGSGPLKVNKSK 60

QY 358 NSKSPPEHLEEMKMWSPNKLHNFHPPKGGPAKPKGHSKPLKAKGRSGILNGQKS 417
Db 61 NSKSPPEHLEEMKMWSPNKLHNFHPPKGGPAKPKGHSKPLKAKGRSGILNGQKS 120

QY 418 TGNSSPKPKGLTPKTKMQLLMDMAKGTQKMTAPRNSGGTPTSSPKHKLPPAALH 477
Db 121 TGNSSPKPKGLTPKTKMQLLMDMAKGTQKMTAPRNSGGTPTSSPKHKLPPAALH 180

QY 478 LIAYYKENDKRDKRSALSCVISTARLLSSDRARLPEELSLVQKRYELLEHKRWAS 537
Db 181 LIAYYKENDKRDKRSALSCVISTARLLSSDRARLPEELSLVQKRYELLEHKRWAS 240

QY 538 MSEEQRKEYLKKRREBKLLKKEKAKERREKEMLEKQYEDQELTGKMLPAFLVD 597
Db 241 MSEEQRKEYLKKRREBKLLKKEKAKERREKEMLEKQYEDQELTGKMLPAFLVD 300

QY 598 TTEGLPNTLFGDVAMVVEFLSCVSGLLLPDAQYPTAVSLMEALSADKGGFLYLNRLVI 657
Db 301 TTEGLPNTLFGDVAMVVEFLSCVSGLLLPDAQYPTAVSLMEALSADKGGFLYLNRLVI 360

QY 658 LIQTLLODRIADYGLGKMLSEIPLTHSVSELVRLCLRRSDVQESGSDTDNDKSA 717
Db 361 LIQTLLODRIADYGLGKMLSEIPLTHSVSELVRLCLRRSDVQESGSDTDNDKSA 420

QY 718 AFEDNEVDQEFLEKLETSFFELTSEKLIQTLALCHRLIMTYSVDHMETQQAELW 777
Db 421 AFEDNEVDQEFLEKLETSFFELTSEKLIQTLALCHRLIMTYSVDHMETQQAELW 480

QY 778 KERLAVLKEENDKRAEKQKREMEAKNKENGKVENGLGKTRKRIIVKFEQVDTEAD 837
Db 481 KERLAVLKEENDKRAEKQKREMEAKNKENGKVENGLGKTRKRIIVKFEQVDTEAD 540

QY 838 MISAVKSRLLAIQAKEREIOEREMKVKL 867
Db 541 MISAVKSRLLAIQAKEREIOEREMKVKL 570

RESULT 15

ABG36635
ID ABG36635 standard; peptide; 572 AA.

AC ABG36635;

DT 19-AUG-2002 (first entry)

DE Human peptide encoded by genome-derived single exon probe SEQ ID 26300.

KW Human; single exon probe; asthma; lung cancer; COPD; ILD;
KW chronic obstructive pulmonary disease; interstitial lung disease;
KW familial idiopathic pulmonary fibrosis; neurofibromatosis;
KW tuberos sclerosis; Gaucher's disease; Niemann-Pick disease;
KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;
KW pulmonary histiocytosis; lymphangioma; fibrocystic pulmonary dysplasia;
KW primary ciliary dyskinesia; pulmonary hypertension;
KW hyaline membrane disease.

OS Homo sapiens.

XX

PN WO200186003-A2.
XX 15-NOV-2001.
XX 30-JAN-2001; 2001WO-US000665.
XX 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2002-114183/15.

DR Spatially-addressable set of single exon nucleic acid probes, used to
PT measure gene expression in human lung samples.

PS Claim 27; SEQ ID NO 26300; 634pp; English.

XX The invention relates to a spatially-addressable set of single exon
CC nucleic acid probes for measuring gene expression in a sample derived
CC from human lung comprising single exon nucleic acid probes having one of
CC 12614 nucleic acid sequences mentioned in the specification, or their
CC complements or the 12387 open reading frames derived from the 12614
CC probes. Also included are a microarray comprising the novel set of probes
CC; the novel set of probes which hybridise at high stringency to a nucleic
CC acid expressed in the human lung; measuring gene expression in a sample
CC derived from human lung, comprising (a) contacting the array with a
CC collection of detectably labeled nucleic acids derived from human lung
CC mRNA, and (b) measuring the label detectably bound to each probe of the
CC array; identifying exons in a eukaryotic genome, comprising (a)
CC algorithmically predicting at least one exon from genomic sequences of
CC the eukaryote; and (b) detecting specific hybridisation of detectably
CC labeled nucleic acids from eukaryote lung mRNA to a single exon probe,
CC having a fragment identical to the predicted exon, the probe is included
CC in the above mentioned microarray; assigning exons to a single gene,
CC comprising (a) identifying exons from genomic sequence by the method
CC above and (b) measuring the expression of each of the exons in several
CC tissues and/or cell types using hybridisation to a single exon
CC microarrays having a probe with the exon, where a common pattern of
CC expression of the exons in the tissues and/or cell types indicates that
CC the exons should be assigned to a single gene; a peptide comprising one
CC of 12011 sequences, mentioned in the specification, or encoded by the
CC probes/open reading frames (ORF). The probes are used for gene expression
CC analysis, and for identifying exons in a gene, particularly using human
CC lung derived mRNA and for the study of lung diseases such as asthma, lung
CC cancer, chronic obstructive pulmonary disease (COPD), interstitial lung
CC disease (ILD), familial idiopathic pulmonary fibrosis, neurofibromatosis,
CC tuberos sclerosis, Gaucher's disease, Niemann-Pick disease, Hermansky-
CC Pudlak syndrome, sarcoidosis, pulmonary haemosiderosis, pulmonary
CC histiocytosis, lymphangioma, fibrocystic pulmonary dysplasia, primary ciliary
CC Karagener syndrome, fibrocystic pulmonary dysplasia, primary ciliary
CC dyskinesia, pulmonary hypertension and hyaline membrane disease. The
CC present sequence is a peptide/protein encoded by a single exon probe of
CC the invention. Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

SQ Sequence 572 AA;

Query Match 36.3%; Score 2889; DB 5; Length 572;

Best Local Similarity 99.5%; Pred. No. 2.5e-185;

Matches 567; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 298 YMTLPSTKRKNTGSPDRKPSKSTDNSSPLNPKLWCHVHLKSLGSGPLKVNKSK 357
Db 1 YMTLPSTKRKNTGSPDRKPSKSTDNSSPLNPKLWCHVHLKSLGSGPLKVNKSK 60

Human; foetal liver; gene expression; single exon nucleic acid probe.
Homo sapiens.
WO200157277-A2.
09-AUG-2001.
30-JAN-2001; 2001WO-US000669.
04-FEB-2000; 2000US-0180312P.
26-MAY-2000; 2000US-0207456P.
30-JUN-2000; 2000US-00608408.
03-AUG-2000; 2000US-00632366.
21-SEP-2000; 2000US-0234687P.
27-SEP-2000; 2000US-0236359P.
04-OCT-2000; 2000GB-00024263.
(MOLE-) MOLECULAR DYNAMICS INC.
Penn SG, Hanzel DK, Chen W, Rank DR;
WPI; 2001-483447/52.
Human genome-derived single exon nucleic acid probes useful for analyzing gene expression in human fetal liver.
Claim 27; SEQ ID NO 29989; 639pp + Sequence Listing; English.
The invention relates to a single exon nucleic acid probe for measuring human gene expression in a sample derived from human foetal liver. The single exon nucleic acid probes may be used for predicting, measuring and displaying gene expression in samples derived from human fetal liver. The present sequence is a peptide encoded by a single exon nucleic acid probe of the invention. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
Sequence 560 AA;
Query Match 35.7%; Score 2847; DB 4; Length 560;
Best Local Similarity 99.8%; Pred. No. 1.6e-182;
Matches 559; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 305 TKRKTGSPDRKPSKSKTDNSLSPLNPKLWCHVHLKSLGSPKVKNSKNSKSP 364
DB 1 TKRKTGSPDRKPSKSKTDNSLSPLNPKLWCHVHLKSLGSPKVKNSKNSKSP 60
QY 365 HLEEMKMSPNKLTNTHFHPKKGPPAKPGKHSDKPLKAKGRSKGILNGQSTGNSKSP 424
DB 61 HLEEMKMSPNKLTNTHFHPKKGPPAKPGKHSDKPLKAKGRSKGILNGQSTGNSKSP 120
QY 425 KGLKTPKTKMOMTLLDMAKCTOKTRAPRNSGGTPTSSKPHKLPALHLIAYKE 484
DB 121 KGLKTPKTKMOMTLLDMAKCTOKTRAPRNSGGTPTSSKPHKLPALHLIAYKE 180
QY 485 NKDREDKRSALSCVISTARLLSSDDRARLPBELSLVOKRYELLEHKKRWASMSSEQRK 544
DB 181 NKDREDKRSALSCVISTARLLSSDDRARLPBELSLVOKRYELLEHKKRWASMSSEQRK 240
QY 545 EYLKXKREELKKLKEKAKERREKEMLERLEQRYEDQELTGKNLPAPFLVDTPEGLPN 604
DB 241 EYLKXKREELKKLKEKAKERREKEMLERLEQRYEDQELTGKNLPAPFLVDTPEGLPN 300
QY 605 TLFGDVAMVVEFLSCYSGLLLPDAQYPITAVSLMEALSADKGGFLYLNRLVILLQTLIQ 664
DB 301 TLFGDVAMVVEFLSCYSGLLLPDAQYPITAVSLMEALSADKGGFLYLNRLVILLQTLIQ 360
QY 665 DRIAEDYGBGLMKLSIPLTHSVSELVRLCLRRSDVQSESGSDTDNKKOSAAAFEDNEV 724
DB 361 DRIAEDYGBGLMKLSIPLTHSVSELVRLCLRRSDVQSESGSDTDNKKOSAAAFEDNEV 420
QY 725 QDEFLEKLETSFELTSEBKLIQILTALCHRLIMTYSVQDHMETRQMSAELWKR L AVL 784

DB 421 QDEFLEKLETSFELTSEBKLIQILTALCHRLIMTYSVQDHMETRQMSAELWKR L AVL 480
QY 785 KEENDKKRAEKOKRKEMEAKNKENGKVENGLCKTDKTKRIVKFEPQVDTAEADMISAVKS 844
DB 481 KEENDKKRAEKOKRKEMEAKNKENGKVENGLCKTDKTKRIVKFEPQVDTAEADMISAVKS 540
QY 845 RLLAIQAKKEREIOEREMK 864
DB 541 RLLAIQAKKEREIOEREMK 560
RESULT 18
AAM30808
ID AAM30808 standard; protein; 560 AA.
XX AC AAM30808;
XX DT 17-OCT-2001 (first entry)
XX DE Peptide #4845 encoded by probe for measuring placental gene expression.
XX KW Probe; microarray; human; placenta; antenatal diagnosis;
XX KW Genetic disorder.
XX OS Homo sapiens.
XX PN WO200157272-A2.
XX PD 09-AUG-2001.
XX 30-JAN-2001; 2001WO-US000663.
XX 04-FEB-2000; 2000US-0180312P.
XX 26-MAY-2000; 2000US-0207456P.
XX 30-JUN-2000; 2000US-00608408.
XX 03-AUG-2000; 2000US-00632366.
XX 21-SEP-2000; 2000US-0234687P.
XX 27-SEP-2000; 2000US-0236359P.
XX 04-OCT-2000; 2000GB-00024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
PA Penn SG, Hanzel DK, Chen W, Rank DR;
PI WPI; 2001-488897/53.
DR Human genome-derived single exon nucleic acid probes useful for analyzing gene expression in human placenta.
PT Claim 27; SEQ ID NO 31077; 654pp; English.
XX The present invention relates to single exon nucleic acid probes (SNP: see AA13135-AA157546). The present sequence is a peptide encoded by one such probe. The probes are useful for producing a microarray for CC predicting, measuring and displaying gene expression in samples derived from human placenta. The probes are useful for antenatal diagnosis of CC human genetic disorders
XX Sequence 560 AA;
Query Match 35.7%; Score 2847; DB 4; Length 560;
Best Local Similarity 99.8%; Pred. No. 1.6e-182;
Matches 559; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 305 TKRKTGSPDRKPSKSKTDNSLSPLNPKLWCHVHLKSLGSPKVKNSKNSKSP 364
DB 1 TKRKTGSPDRKPSKSKTDNSLSPLNPKLWCHVHLKSLGSPKVKNSKNSKSP 60
QY 365 HLEEMKMSPNKLTNTHFHPKKGPPAKPGKHSDKPLKAKGRSKGILNGQSTGNSKSP 424
DB 61 HLEEMKMSPNKLTNTHFHPKKGPPAKPGKHSDKPLKAKGRSKGILNGQSTGNSKSP 120

QY 425 KKGLKTPKTKMKQMTLLDWAQTKQWTRAPRNSGGTPTRTSSKPKHLPAAHLIAYYKE 484
DB 121 KKGLKTPKTKMKQMTLLDWAQTKQWTRAPRNSGGTPTRTSSKPKHLPAAHLIAYYKE 180
QY 485 NKREDKRSALSCVISTKARLLSSSDRARLPPEELRSLVQKRYELLEHKRWASMESEQRK 544
DB 181 NKREDKRSALSCVISTKARLLSSSDRARLPPEELRSLVQKRYELLEHKRWASMESEQRK 240
QY 545 EYLKRRBELKKLKEKAKERREKEMLEKQRYEDELTKGNLPAPFLVDTPEGLPN 604
DB 241 EYLKRRBELKKLKEKAKERREKEMLEKQRYEDELTKGNLPAPFLVDTPEGLPN 300
QY 605 TLFGDVAMVVEFLSCYSGLLLPDAQYPITAVSLMEALSADKGGFLYLNRLVILLQTLQ 664
DB 301 TLFGDVAMVVEFLSCYSGLLLPDAQYPITAVSLMEALSADKGGFLYLNRLVILLQTLQ 360
QY 665 DEIAEDYGELGKMLSEIPLTLHSVSELVRLCLRRSDVOEESGSDTDNKSAAFPEDNEV 724
DB 361 DEIAEDYGELGKMLSEIPLTLHSVSELVRLCLRRSDVOEESGSDTDNKSAAFPEDNEV 420
QY 725 QDEFLEKLETSEFFELTSEEKLIQILTALCHRLMTYSVQDHMETRQQMSAELWKERLAVL 784
DB 421 QDEFLEKLETSEFFELTSEEKLIQILTALCHRLMTYSVQDHMETRQQMSAELWKERLAVL 480
QY 785 KEENDKKRAEQKREKEMAKNKENGKVENGLGKTDKRRKRIVKFEPQVDTAEADMISAVKS 844
DB 481 KEENDKKRAEQKREKEMAKNKENGKVENGLGKTDKRRKRIVKFEPQVDTAEADMISAVKS 540
QY 845 RLLAIQAQKEREIQEREMK 864
DB 541 RLLAIQAQKEREIQEREMK 560
RESULT 19
ABB32101
ID ABB32101 standard; peptide; 560 AA.
XX
AC ABB32101;
XX
DT 01-FEB-2002 (first entry)
XX
DE Peptide #4752 encoded by breast cell single exon nucleic acid probe.
XX
KW Human; microarray; single exon probe; gene expression; breast; disease;
cancer.
XX
OS Homo sapiens.
XX
PN WO200157271-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US000662.
XX
PR 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-496933/54.
XX
DR
XX
PT New spatially-addressable set of single exon nucleic acid probes, useful
PT for measuring gene expression in sample derived from human breast,
PT comprises number of single exon nucleic acid probes.
XX
XX
PS Claim 27; SEQ ID NO 15069; 327pp + Sequence Listing; English.

XX The invention relates to a spatially-addressable set of single exon
CC nucleic acid probes for measuring gene expression in a sample derived
CC from human breast and Br 474 cells. The method involves contacting the
CC probes with a collection of detectably labelled nucleic acids derived
CC from mRNA of human breast, and then measuring the label bound to each
CC probe of the microarray. The probes are useful for verifying the
CC expression of regions of genomic DNA predicted to encode proteins. They
CC are useful for gene discovery, and for determining predisposition and/or
CC prognosing breast disease. Gene expression analysis is useful for
CC assessing the toxicity of chemical agents on cells. The microarray of
CC this invention presents a far greater diversity of probes for measuring
CC gene expression, with far less bias than expressed sequence tag
CC microarrays. The method is suitable for rapid production of functional
CC information from genomic sequence. The present sequence is a peptide
CC encoded by a single exon nucleic acid probe of the invention. Note: The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 560 AA;
Query Match 35.7%; Score 2847; DB 4; Length 560;
Best Local Similarity 99.8%; Pred. No. 1.6e-182;
Matches 559; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 305 TKRNTGSPDRKPSKSKTNDSSLSPLNPKLWCHVHLKKSLSGSPLVKVKSKNSKSP 364
DB 1 TKRNTGSPDRKPSKSKTNDSSLSPLNPKLWCHVHLKKSLSGSPLVKVKSKNSKSP 60
QY 365 HLEEMKMSPNKLTHTNPHIPKGPAPKPKGKSDKPLKAGRSKGIILNGKSTGNSKSP 424
DB 61 HLEEMKMSPNKLTHTNPHIPKGPAPKPKGKSDKPLKAGRSKGIILNGKSTGNSKSP 120
QY 425 KKGLKTPKTKMKQMTLLDWAQTKQWTRAPRNSGGTPTRTSSKPKHLPAAHLIAYYKE 484
DB 121 KKGLKTPKTKMKQMTLLDWAQTKQWTRAPRNSGGTPTRTSSKPKHLPAAHLIAYYKE 180
QY 485 NKREDKRSALSCVISTKARLLSSSDRARLPPEELRSLVQKRYELLEHKRWASMESEQRK 544
DB 181 NKREDKRSALSCVISTKARLLSSSDRARLPPEELRSLVQKRYELLEHKRWASMESEQRK 240
QY 545 EYLKRRBELKKLKEKAKERREKEMLEKQRYEDELTKGNLPAPFLVDTPEGLPN 604
DB 241 EYLKRRBELKKLKEKAKERREKEMLEKQRYEDELTKGNLPAPFLVDTPEGLPN 300
QY 605 TLFGDVAMVVEFLSCYSGLLLPDAQYPITAVSLMEALSADKGGFLYLNRLVILLQTLQ 664
DB 301 TLFGDVAMVVEFLSCYSGLLLPDAQYPITAVSLMEALSADKGGFLYLNRLVILLQTLQ 360
QY 665 DEIAEDYGELGKMLSEIPLTLHSVSELVRLCLRRSDVOEESGSDTDNKSAAFPEDNEV 724
DB 361 DEIAEDYGELGKMLSEIPLTLHSVSELVRLCLRRSDVOEESGSDTDNKSAAFPEDNEV 420
QY 725 QDEFLEKLETSEFFELTSEEKLIQILTALCHRLMTYSVQDHMETRQQMSAELWKERLAVL 784
DB 421 QDEFLEKLETSEFFELTSEEKLIQILTALCHRLMTYSVQDHMETRQQMSAELWKERLAVL 480
QY 785 KEENDKKRAEQKREKEMAKNKENGKVENGLGKTDKRRKRIVKFEPQVDTAEADMISAVKS 844
DB 481 KEENDKKRAEQKREKEMAKNKENGKVENGLGKTDKRRKRIVKFEPQVDTAEADMISAVKS 540
QY 845 RLLAIQAQKEREIQEREMK 864
DB 541 RLLAIQAQKEREIQEREMK 560
RESULT 20
ABB22640
ID ABB22640 standard; protein; 560 AA.
XX
XX ABB22640;
XX

DT 23-JAN-2002 (first entry)
XX Protein #4639 encoded by probe for measuring heart cell gene expression.
DE Human; gene expression; heart; microarray; vascular system;
KW cardiovascular disease; hypertension; cardiac arrhythmia;
KW congenital heart disease.
XX Homo sapiens.
XX WO200157274-A2.
XX 09-AUG-2001.
XX 30-JAN-2001; 2001WO-US000666.
XX 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-48899/53.
XX Single exon nucleic acid probes for analyzing gene expression in human
PT hearts.
XX Claim 15; SEQ ID NO 24410; 530pp; English.
XX The present invention relates to single exon nucleic acid probes for
CC measuring human gene expression in a sample derived from human heart (see
CC ABA21535-ABA41305). The present sequence is a protein encoded by one such
CC probe. The probes may be used for predicting, measuring and displaying
CC gene expression in samples derived from the human heart via microarrays.
CC By measuring gene expression, the probes are useful for predicting
CC diagnosing, grading, staging, monitoring and prognosing diseases of the
CC human heart and vascular system e.g. cardiovascular disease,
CC hypertension, cardiac arrhythmias and congenital heart disease. Note: The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences
XX Sequence 560 AA;
XX
XX Query Match 35.7%; Score 2847; DB 4; Length 560;
XX Best Local Similarity 99.8%; Pred. No. 1.6e-182;
XX Matches 559; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 305 TTKRNTGSPDRKPKSKSKTDSLSPLNPKLWCHVHLKSLGSPKLVKNKSKSPPE 364
Db 1 TTKRNTGSPDRKPKSKSKTDSLSPLNPKLWCHVHLKSLGSPKLVKNKSKSPPE 60
QY 365 HLEEMMKMSPNKLHNFHPPKGPAPKYPGKHSDKPLKAGRSKGLNGQSTGNSKSP 424
Db 61 HLEEMMKMSPNKLHNFHPPKGPAPKYPGKHSDKPLKAGRSKGLNGQSTGNSKSP 120
QY 425 KXGLTKPTKMKQMTLLDMAKGTQKMTAPRNSGGTPTSSKPHKLPALHLIAYYKE 484
Db 121 KXGLTKPTKMKQMTLLDMAKGTQKMTAPRNSGGTPTSSKPHKLPALHLIAYYKE 180
QY 485 NKDREDKRSALSCVISKTLARLSSSEDARLPEELRSVLQRYELLEHKKRWASMSBEOQRK 544
Db 181 NKDREDKRSALSCVISKTLARLSSSEDARLPEELRSVLQRYELLEHKKRWASMSBEOQRK 240
QY 545 EYLKXKREBELKKLKEKAKEREKEMLEKRYEDQELTGKNLPAPRLVDTPEGLPN 604
Db 241 EYLKXKREBELKKLKEKAKEREKEMLEKRYEDQELTGKNLPAPRLVDTPEGLPN 300

QY 605 TLFQGVAMVVEFLSCYSGLLLLPDAQYPTAVSLMEALSADKGGFLYNRVLVILLQTLQ 664
Db 301 TLFQGVAMVVEFLSCYSGLLLLPDAQYPTAVSLMEALSADKGGFLYNRVLVILLQTLQ 360
QY 665 DEIAEDYGELGKMLSEIPLTLHVSVELVRLCLRRSDVQEESESGSDTDDNKDSAAAFEDNEV 724
Db 361 DEIAEDYGELGKMLSEIPLTLHVSVELVRLCLRRSDVQEESESGSDTDDNKDSAAAFEDNEV 420
QY 725 QDEFLEKLETSEFFELTSEELQIILTALCHRLIMTYSVQDHMETROQMSAELWKERLAVL 784
Db 421 QDEFLEKLETSEFFELTSEELQIILTALCHRLIMTYSVQDHMETROQMSAELWKERLAVL 480
QY 785 KEENDKGAERKOKRMEAKNKENGKVENGLKTRKRIIVKPEPOVDTEADMISAVKS 844
Db 481 KEENDKGAERKOKRMEAKNKENGKVENGLKTRKRIIVKPEPOVDTEADMISAVKS 540
QY 845 RELALAIQAKKEREIOEREMK 864
Db 541 RELALAIQAKKEREIOEREMK 560
RESULT 21
AAM70484
ID AAM70484 standard; protein; 560 AA.
XX AC AAM70484;
XX AC AAM70484;
DT 06-NOV-2001 (first entry)
XX DE Human bone marrow expressed probe encoded protein SEQ ID NO: 30790.
XX KW Human; bone marrow expressed exon; gene expression analysis; probe;
XX KW microarray; cancer; leukaemia; lymphoma; myeloma.
XX OS Homo sapiens.
XX PN WO200157276-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US000668.
PR 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-48899/53.
XX Human genome-derived single exon nucleic acid probes useful for analyzing
PT gene expression in human bone marrow.
XX Example 4; SEQ ID NO 30790; 658pp + Sequence Listing; English.
XX The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC bone marrow. They can be used to measure gene expression in bone marrow
CC samples, which may enable the improved diagnosis and treatment of cancers
CC such as lymphoma, leukaemia and myeloma. The present sequence is a
CC protein encoded by one of the probes of the invention
XX Sequence 560 AA;
SQ
Query Match 35.7%; Score 2847; DB 4; Length 560;
Best Local Similarity 99.8%; Pred. No. 1.6e-182;

Matches 559; Conservative 0; Mismatches 1; Indels 0; Gaps 0;	
QY	305 TKRNTGSPDRPKSKSTDNSSLSPLNPKLWCHVHLKKSLSGSPLVKVKNSKNSKSP 364
DB	1 TKRNTGSPDRPKSKSKTDNSSLSPLNPKLWCHVHLKKSLSGSPLVKVKNSKNSKSP 60
QY	365 HLEEMKQWSPNKLHTNFHPIPKGPPAKPKGKHSKPLKAKGRSGKILNGQSTGNSKSP 424
DB	61 HLEEMKQWSPNKLHTNFHPIPKGPPAKPKGKHSKPLKAKGRSGKILNGQSTGNSKSP 120
QY	425 KKGKLTPTKTKQMTLLDMAKTQMTAPRNSGGTPTTSSKPHKHLPPAALHLIAYKE 484
DB	121 KKGKLTPTKTKQMTLLDMAKTQMTAPRNSGGTPTTSSKPHKHLPPAALHLIAYKE 180
QY	485 NKREDKRSALSCVISTARLLSSDRARLPBELRSVLQKRYELLEHKKRWASMESEOR 544
DB	181 NKREDKRSALSCVISTARLLSSDRARLPBELRSVLQKRYELLEHKKRWASMESEOR 240
QY	545 EYLKKEELKKLKEKAKERREKEMLERLEKQRYEDQELTGKNLPAFRLVDTPEGLPN 604
DB	241 EYLKKEELKKLKEKAKERREKEMLERLEKQRYEDQELTGKNLPAFRLVDTPEGLPN 300
QY	605 TLFGDVAMVVEFLSCYSGILLPDAQYPTAVSLMEALSADKGGFLYLNRLVILLQTL 664
DB	301 TLFGDVAMVVEFLSCYSGILLPDAQYPTAVSLMEALSADKGGFLYLNRLVILLQTL 360
QY	665 DEIAEDYGELGKMLSEIPLTHSVSELVRLCLRRSDVQESGSDTDNKSAAAFEDNEV 724
DB	361 DEIAEDYGELGKMLSEIPLTHSVSELVRLCLRRSDVQESGSDTDNKSAAAFEDNEV 420
QY	725 QDEFLEKLTSEFFELTSEEKLIQILTALCHRLIMTYSVDHMETRQQMSAELWKERLAV 784
DB	421 QDEFLEKLTSEFFELTSEEKLIQILTALCHRLIMTYSVDHMETRQQMSAELWKERLAV 480
QY	785 KEENDKRAEQKQKMEAKNKENGKVENGLGKTDKRRKRIVKFEPQVDTAEEDMISAVKS 844
DB	481 KEENDKRAEQKQKMEAKNKENGKVENGLGKTDKRRKRIVKFEPQVDTAEEDMISAVKS 540
QY	845 RLLAIAQAKKEREIOEREMK 864
DB	541 RLLAIAQAKKEREIOEREMK 560
RESULT 22	
AAM58044	
ID	AAM58044 standard; protein; 560 AA.
XX	
AC	AAM58044;
DT	
DT	05-NOV-2001 (first entry)
XX	
DE	Human brain expressed single exon probe encoded protein SEQ ID NO: 30149.
XX	
KW	Human; brain expressed exon; gene expression analysis; probe; microarray;
KW	Alzheimer's disease; multiple sclerosis; schizophrenia; epilepsy; cancer.
XX	
OS	Homo sapiens.
XX	
PN	WO200157275-A2.
XX	
PD	09-AUG-2001.
XX	
PF	30-JAN-2001; 2001WO-US0000667.
XX	
PR	04-FEB-2000; 2000US-0180312P.
PR	26-MAY-2000; 2000US-0207456P.
PR	30-JUN-2000; 2000US-00608408.
PR	03-AUG-2000; 2000US-00632366.
PR	21-SEP-2000; 2000US-0234687F.
PR	27-SEP-2000; 2000US-0236359P.
PR	04-OCT-2000; 2000GB-00024263.
XX	
PA	(MOLE-) MOLECULAR DYNAMICS INC.
Penn SG, Hanzel DK, Chen W, Rank DR;	
WPI; 2001-483446/52.	
Single exon nucleic acid probes for analyzing gene expression in human brains.	
Example 4; SEQ ID NO 30149; 650pp + Sequence Listing; English.	
The present invention provides a number of single exon nucleic acid probes which are derived from genomic sequences expressed in the human brain. They can be used to measure gene expression in brain cell samples, which may enable the diagnosis and improved treatment of nervous system diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia, epilepsy and cancers. The present sequence is a protein encoded by one of the probes of the invention	
Sequence 560 AA;	
Query Match 35.7%; Score 2847; DB 4; Length 560;	
Best Local Similarity 99.8%; Pred. No. 1.6e-182;	
Matches 559; Conservative 0; Mismatches 1; Indels 0; Gaps 0;	
QY	305 TKRNTGSPDRPKSKSKTDNSSLSPLNPKLWCHVHLKKSLSGSPLVKVKNSKNSKSP 364
DB	1 TKRNTGSPDRPKSKSKTDNSSLSPLNPKLWCHVHLKKSLSGSPLVKVKNSKNSKSP 60
QY	365 HLEEMKQWSPNKLHTNFHPIPKGPPAKPKGKHSKPLKAKGRSGKILNGQSTGNSKSP 424
DB	61 HLEEMKQWSPNKLHTNFHPIPKGPPAKPKGKHSKPLKAKGRSGKILNGQSTGNSKSP 120
QY	425 KKGKLTPTKTKQMTLLDMAKTQMTAPRNSGGTPTTSSKPHKHLPPAALHLIAYKE 484
DB	121 KKGKLTPTKTKQMTLLDMAKTQMTAPRNSGGTPTTSSKPHKHLPPAALHLIAYKE 180
QY	485 NKREDKRSALSCVISTARLLSSDRARLPBELRSVLQKRYELLEHKKRWASMESEOR 544
DB	181 NKREDKRSALSCVISTARLLSSDRARLPBELRSVLQKRYELLEHKKRWASMESEOR 240
QY	545 EYLKKEELKKLKEKAKERREKEMLERLEKQRYEDQELTGKNLPAFRLVDTPEGLPN 604
DB	241 EYLKKEELKKLKEKAKERREKEMLERLEKQRYEDQELTGKNLPAFRLVDTPEGLPN 300
QY	605 TLFGDVAMVVEFLSCYSGILLPDAQYPTAVSLMEALSADKGGFLYLNRLVILLQTL 664
DB	301 TLFGDVAMVVEFLSCYSGILLPDAQYPTAVSLMEALSADKGGFLYLNRLVILLQTL 360
QY	665 DEIAEDYGELGKMLSEIPLTHSVSELVRLCLRRSDVQESGSDTDNKSAAAFEDNEV 724
DB	361 DEIAEDYGELGKMLSEIPLTHSVSELVRLCLRRSDVQESGSDTDNKSAAAFEDNEV 420
QY	725 QDEFLEKLTSEFFELTSEEKLIQILTALCHRLIMTYSVDHMETRQQMSAELWKERLAV 784
DB	421 QDEFLEKLTSEFFELTSEEKLIQILTALCHRLIMTYSVDHMETRQQMSAELWKERLAV 480
QY	785 KEENDKRAEQKQKMEAKNKENGKVENGLGKTDKRRKRIVKFEPQVDTAEEDMISAVKS 844
DB	481 KEENDKRAEQKQKMEAKNKENGKVENGLGKTDKRRKRIVKFEPQVDTAEEDMISAVKS 540
QY	845 RLLAIAQAKKEREIOEREMK 864
DB	541 RLLAIAQAKKEREIOEREMK 560
RESULT 23	
AAM05928	
ID	AAM05928 standard; protein; 560 AA.
XX	
AC	AAM05928;
DT	
DT	09-OCT-2001 (first entry)
XX	

DE Peptide #4610 encoded by probe for measuring breast gene expression.
 XX Probe; human; breast disease; breast cancer; development disorder;
 KW inflammatory disease; proliferative breast disease; non-carcinoma tumour.
 XX Homo sapiens.
 OS
 XX
 FN WO200157270-A2.
 XX
 PD 09-AUG-2001.
 XX
 XX 29-JAN-2001; 2001WO-US000661.
 XX
 XX 04-FEB-2000; 2000US-0180312P.
 PR 26-MAY-2000; 2000US-0207456P.
 PR 30-JUN-2000; 2000US-00608408.
 PR 03-AUG-2000; 2000US-00632366.
 PR 21-SEP-2000; 2000US-0234687P.
 PR 27-SEP-2000; 2000US-0236359P.
 PR 04-OCT-2000; 2000GB-00024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 XX Penn SG, Hanzel DK, Chen W, Rank DR;
 PI
 XX WPI; 2001-476286/51.
 XX
 XX Novel single exon nucleic acid probe used to measuring gene expression in
 PT a human breast.
 PT
 XX Claim 27; SEQ ID NO 14668; 322pp; English.
 XX
 XX The present invention relates to novel single exon nucleic acid probes
 CC (see AAT00010-AA110067). The present sequence is a peptide encoded by one
 CC such probe. The probes are useful for measuring human gene expression in
 CC a human breast sample, where the probe hybridises at high stringency to a
 CC nucleic acid expressed in the human breast. The probes are useful for
 CC predicting, diagnosing, grading, staging, monitoring and prognosing
 CC diseases of the human breast, particularly those diseases with polygenic
 CC aetiology. The diseases include: breast cancer; disorders of development,
 CC inflammatory diseases of the breast; fibrocystic changes, proliferative
 CC breast disease and non-carcinoma tumours. Note: The sequence data for
 CC this patent did not form part of the printed specification, but was
 CC obtained in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 560 AA;

Query Match 35.7%; Score 2847; DB 4; Length 560;
 Best Local Similarity 99.8%; Pred. No. 1.6e-182;
 Matches 559; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 305 TTKKNTGSDRPSKSKTDNSLSPLNPKLVCHVHLKKSLSGSPKLVNKNKSPBE 364
 DB 1 TTKKNTGSDRPSKSKTDNSLSPLNPKLVCHVHLKKSLSGSPKLVNKNKSPBE 60
 QY 365 HLEEMKMSPNKLTNTHFHPKGGPAKPGKHSKPLKAKGRSGILNGQSTGNSKP 424
 DB 61 HLEEMKMSPNKLTNTHFHPKGGPAKPGKHSKPLKAKGRSGILNGQSTGNSKP 120
 QY 425 KKGLKTPKTKMQMTLLDMAKTQKMTAPRNSGGTPTSSKPHKLPALHLYAYKE 484
 DB 121 KKGLKTPKTKMQMTLLDMAKTQKMTAPRNSGGTPTSSKPHKLPALHLYAYKE 180
 QY 485 NKDREDKRSALSCVSKTARLLSSEDRARLPBELSLVQRYELLEHKRWASMSSEQRK 544
 DB 181 NKDREDKRSALSCVSKTARLLSSEDRARLPBELSLVQRYELLEHKRWASMSSEQRK 240
 QY 545 EYLKKREELKKLKEKAKEREKEMLELEKQRYEDELTKNLPAPRLVDTPEGLPN 604
 DB 241 EYLKKREELKKLKEKAKEREKEMLELEKQRYEDELTKNLPAPRLVDTPEGLPN 300
 QY 605 TLFGDVAMVVEFLSCYSGLLPDAQVPITAVSLMEALSADKGGFLYLNRLVILLQTLQ 664

DB 301 TLFGDVAMVVEFLSCYSGLLPDAQVPITAVSLMEALSADKGGFLYLNRLVILLQTLQ 360
 QY 665 DETAEDYGLGKWLSEIPLTLHSVSELVRLCYRRSDVQSESGSDTDNKSAAAFEDNEV 724
 DB 361 DETAEDYGLGKWLSEIPLTLHSVSELVRLCYRRSDVQSESGSDTDNKSAAAFEDNEV 420
 QY 725 QDEFLEKLETSFFELTSEKLIQTALCHRLIMTYSVQDHMETROQMSAELWKERLAVL 784
 DB 421 QDEFLEKLETSFFELTSEKLIQTALCHRLIMTYSVQDHMETROQMSAELWKERLAVL 480
 QY 785 KEENDKRAEKOKRKEAMEAKNKENGKVENGLGKTRKRIIVKPEPQVDTAEADMISAVKS 844
 DB 481 KEENDKRAEKOKRKEAMEAKNKENGKVENGLGKTRKRIIVKPEPQVDTAEADMISAVKS 540
 QY 845 RLLAIQAKKEREIOEREMK 864
 DB 541 RLLAIQAKKEREIOEREMK 560

RESULT 24

ABG40123
 ID ABG40123 standard; peptide; 560 AA.
 AC
 XX ABG40123;
 XX
 DT 19-AUG-2002 (first entry)
 XX
 XX Human peptide encoded by genome-derived single exon probe SEQ ID 29788.
 DE
 XX Human; single exon probe; asthma; lung cancer; COPD; ILD;
 KW chronic obstructive pulmonary disease; interstitial lung disease;
 KW familial idiopathic pulmonary fibrosis; neurofibromatosis;
 KW tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;
 KW Hermansky-Rudrak syndrome; sarcoidosis; pulmonary haemosiderosis;
 KW pulmonary histiocytosis; lymphangioleiomyomatosis; Karagener syndrome;
 KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
 KW primary ciliary dyskinesia; pulmonary hypertension;
 KW hyaline membrane disease.
 XX
 XX Homo sapiens.
 OS
 XX WO200186003-A2.
 PN
 XX 15-NOV-2001.
 PD
 XX 30-JAN-2001; 2001WO-US000665.
 PF
 XX 04-FEB-2000; 2000US-0180312P.
 PR 26-MAY-2000; 2000US-0207456P.
 PR 30-JUN-2000; 2000US-00608408.
 PR 03-AUG-2000; 2000US-00632366.
 PR 21-SEP-2000; 2000US-0234687P.
 PR 27-SEP-2000; 2000US-0236359P.
 PR 04-OCT-2000; 2000GB-00024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 XX Penn SG, Hanzel DK, Chen W, Rank DR;
 PI
 XX WPI; 2002-114183/15.
 XX
 XX Spatially-addressable set of single exon nucleic acid probes, used to
 PT measure gene expression in human lung samples.
 PT
 XX Claim 27; SEQ ID NO 29788; 634pp; English.
 XX
 XX The invention relates to a spatially-addressable set of single exon
 CC nucleic acid probes for measuring gene expression in a sample derived
 CC from human lung comprising single exon nucleic acid probes having one of
 CC 12614 nucleic acid sequences mentioned in the specification, or their
 CC complements or the 12387 open reading frames derived from the 12614
 CC probes. Also included are a microarray comprising the novel set of probes

CC ; the novel set of probes which hybridise at high stringency to a nucleic
CC acid expressed in the human lung; measuring gene expression in a sample
CC derived from human lung, comprising (a) contacting the array with a
CC collection of detectably labeled nucleic acids derived from human lung
CC mRNA, and (b) measuring the label detectably bound to each probe of the
CC array; identifying exons in a eukaryotic genome, comprising (a)
CC algorithmically predicting at least one exon from genomic sequences of
CC the eukaryote; and (b) detecting specific hybridisation of detectably
CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,
CC having a fragment identical to the predicted exon, the probe is included
CC in the above mentioned microarray; assigning exons to a single gene,
CC comprising (a) identifying exons from genomic sequence by the method
CC above and (b) measuring the expression of each of the exons in several
CC tissues and/or cell types using hybridisation to a single exon
CC microarrays having a probe with the exon, where a common pattern of
CC expression of the exons in the tissues and/or cell types indicates that
CC the exons should be assigned to a single gene; a peptide comprising one
CC of 12011 sequences, mentioned in the specification, or encoded by the
CC probes/open reading frames (ORF). The probes are used for gene expression
CC analysis, and for identifying exons in a gene, particularly using human
CC lung derived mRNA and for the study of lung diseases such as asthma, lung
CC cancer, chronic obstructive pulmonary disease (COPD), interstitial lung
CC disease (ILD), familial idiopathic pulmonary fibrosis, neurofibromatosis,
CC tuberous sclerosis, Gaucher's disease, Niemann-Pick disease, Hermansky-
CC Pudlak syndrome, sarcoidosis, pulmonary haemosiderosis, pulmonary
CC histiocytosis, lymphangioleiomyomatosis, pulmonary alveolar proteinosis,
CC Karsagen syndrome, fibrocystic pulmonary dysplasia, primary ciliary
CC dyskinesia, pulmonary hypertension and hyaline membrane disease. The
CC present sequence is a peptide/protein encoded by a single exon probe of
CC the invention. Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX Sequence 560 AA;

Query Match 35.7%; Score 2847; DB 5; Length 560;
Best Local Similarity 99.8%; Pred. No. 1.6e-182;
Matches 559; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 305 TTKKNTGSPDRKPSKSKTDSNLSPLNPKLWCHVHLKKSLSGSPKVKNSKNSKSP 364
DB 1 TTKKNTGSPDRKPSKSKTDSNLSPLNPKLWCHVHLKKSLSGSPKVKNSKNSKSP 60
QY 365 HLEEMKMKSPNKLHTNFI PKKGPAPKPGKHSKPKLAKGRSGILNGKSTGNSKP 424
DB 61 HLEEMKMKSPNKLHTNFI PKKGPAPKPGKHSKPKLAKGRSGILNGKSTGNSKP 120
QY 425 KKGLKTPKTKMKQMTLLDMAKGTQKWRAPRNSGGTPTRTSSKPKHLPPAALHLIAYYKE 484
DB 121 KKGLKTPKTKMKQMTLLDMAKGTQKWRAPRNSGGTPTRTSSKPKHLPPAALHLIAYYKE 180
QY 485 NKDREDKRSALSCVSKTARLLSSBDRAPLPELRSVQKRYELHKKRWASMSGEOR 544
DB 181 NKDREDKRSALSCVSKTARLLSSBDRAPLPELRSVQKRYELHKKRWASMSGEOR 240
QY 545 EYLKKREELKKLKEKAKERKEKEMLEKQKYEDELTKNLPAPRLVDTPEGLPN 604
DB 241 EYLKKREELKKLKEKAKERKEKEMLEKQKYEDELTKNLPAPRLVDTPEGLPN 300
QY 605 TLFGDVAMVVEFLSCVSGLLLPDAQYPTAVSLMEALSADGGFLYLNRLVLIQTLLQ 664
DB 301 TLFGDVAMVVEFLSCVSGLLLPDAQYPTAVSLMEALSADGGFLYLNRLVLIQTLLQ 360
QY 665 DEIAEDYGBGLMKLSEIPLTHSVSELVRLCLRRSDVQESGSDTDNKKSAAFEDNEV 724
DB 361 DEIAEDYGBGLMKLSEIPLTHSVSELVRLCLRRSDVQESGSDTDNKKSAAFEDNEV 420
QY 725 QDEFLKLTSEFFELTSEKLIQILTALCHRLTMTYSVODHMETQOQMSAELWKLRLAVL 784
DB 421 QDEFLKLTSEFFELTSEKLIQILTALCHRLTMTYSVODHMETQOQMSAELWKLRLAVL 480
QY 785 KEENDKKRAEKQKREMEAKNKENGKVENGLGKTRKKRIKVFEPQVDTAEADMISAVKS 844

DB 481 KEENDKKRAEKQKREMEAKNKENGKVENGLGKTRKKRIKVFEPQVDTAEADMISAVKS 540
QY 845 RRLIAIOAKKEREIOEREMK 864
DB 541 RRLIAIOAKKEREIOEREMK 560
RESULT 25
ID AAM00760 standard; protein; 513 AA.
XX AAM00760;
AC AAM00760;
DT 01-OCT-2001 (first entry)
XX Human bone marrow protein, SEQ ID NO: 123.
DE Human; bone marrow; antiinflammatory; cytostatic; neuroprotective;
KW antiviral; antibacterial; antifungal; anti-HIV; haemostatic;
KW immunosuppressive; gene therapy; cytokine cell proliferation;
KW cell differentiation modulator; immune disorder; infection; cancer;
KW human immunodeficiency virus; HIV; autoimmune disorder; haemophilia.
XX Homo sapiens.
OS WO200153453-A2.
PN 26-JUL-2001.
PD 23-DEC-2000; 2000WO-US034960.
XX 23-DEC-1999; 99US-00471275.
PR 21-JAN-2000; 2000US-00488725.
PR 25-APR-2000; 2000US-00552317.
PR 20-JUN-2000; 2000US-00598042.
PR 19-JUL-2000; 2000US-00620312.
PR 03-AUG-2000; 2000US-00653450.
PR 14-SEP-2000; 2000US-00662191.
PR 19-OCT-2000; 2000US-00693036.
PR 30-NOV-2000; 2000US-0250583P.
XX (HYSE-) HYSEQ INC.

XX Ford JE, Boyle BJ, Tang YT, Liu C, Asundi V, Chen R, Ma Y;
PI Ren F, Wang J, Werhman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;
PI Zhou P, Drmanac RI;
XX WPI; 2001-488707/53.
DR N-PSDB; AAH89879.
XX Novel bone-marrow-expressed polynucleotides and polypeptides, useful for
PT treating e.g. cancer and immune deficiency disorders.
XX Claim 10; Page 292-293; 648pp; English.
XX The present sequence is one of 251 novel human polypeptides encoded by a
CC bone marrow-expressed polynucleotide. The polynucleotide and the
CC polypeptide encoded by it are useful in the treatment of various immune
CC deficiencies and disorders. The deficiencies and disorders may be
CC genetic, may be caused by a viral (e.g. HIV), bacterial or fungal
CC infection, or may result from an autoimmune disorder, a coagulation
CC disorder (e.g. haemophilia), inhibition of tumour cell proliferation,
CC suppression of an inflammatory response or treatment of a nervous system
CC disorder such as Alzheimer's disease. Detection of the presence or
CC increased expression of the polynucleotide or the protein it encodes is
CC useful for the diagnosis and/or prognosis of one or more types of cancer.
CC The polynucleotide and polypeptide can be used as nutritional sources or
CC supplements and in the screening of chemical compounds as potential drugs
XX Sequence 513 AA;

Query Match 28.9%; Score 2305.5; DB 4; Length 513;
Best Local Similarity 63.5%; Pred. No. 3.6e-146;

Matches	477;	Conservative	5;	Mismatches	8;	Indels	261;	Gaps	7;
QY	758	MTYSVDHMETRQMSAEIWKERLAVLKEENDKRAEKQKREMEAKNKENGKVENGLGK	817						
Db	1	MTYSVDHMETRQMSAEIWKERLAVLKEENDKRAEKQKREMEAKNKENGKVENGLGK	60						
QY	818	TDRKKELVKEPQVDTEAEDMTSAVKSRLLAIOAKKEREIOEREMKVKLERQAEERTR	877						
Db	61	TDRKKELVKEPQVDTEAEDMTSAVKSRLLAIOAKKEREIOEREMK	107						
QY	878	KHKAAEKAQEGIAKAKLVMRRTPGTDRNHNRYWLFSDVEVGLPIEKGVHDSIDYRF	937						
Db	108	-----ELSGLP-----	113						
QY	938	NHHCKDHTVSGDEDYCPRSKKNAGNKASMTQHGTADEVAVETTTPKQGNLWFLCDSQ	997						
Db	114	-----CPYR-----	123						
QY	998	KELDELLNCLHPQIGRESQIKERLEKRYQDIHSHLARKPNLGLKSCDGNQELLNFLRS	1057						
Db	124	KELDELLNCLHPQIGRESQIKERLEKRYQDIHSHLARKPNLGLKSCDGNQELLNFLRS	183						
QY	1058	DLIEVATRLQKGLGVVEETSEFEARVISLEKLKDFGECVIALQASVIKKFLQGFMAPKQ	1117						
Db	184	DLIEVATRLQKGLGVVEETSEFEAR	209						
QY	1118	KRRKLQSDSAKTEEVDEEKKMVEAKVASALEKWKTAIREAQTFSRMHVLGMLDACIK	1177						
Db	210	-----VASALEKWKTAIREAQTFSRMHVLGMLDACIK	242						
QY	1178	WDMASENARCKVCPKGGEDDKLILDCENKAFLFCRLPALYEVDPGEWQCACOPATAR	1237						
Db	243	WDMASENARCKVCPKGGEDDKLILDCENKAFLFCRLPALYEVDP	288						
QY	1238	RNSRGRNYTEESASEDSEDESEEEEEDEVEAGLRPRKTIIRGKHSVIPP	1297						
Db	289	-----VAPRKTIIRGKHSVIPP	304						
QY	1298	AARSGRRPGKPHSTRSQKAPPVDDAEVDELVLQTKRSSRQSLQKCEIHLKVIK	1357						
Db	305	AARSGRRPGKPHSTRSQKAPPVDDAEVDELVLQTKRSSRQSLQKCEIHLKVIK	364						
QY	1358	YFSPWPF-----REPVTDEADYDVITHPMDQTVQNKCS	1394						
Db	365	YFSPWPFCTLSGRGTAVKAVQILHLVLHREBPVTRDEADYDVITHPMDQTVQNKCS	424						
QY	1395	CGSYRSVQBELTDMKQVFTNAEYVNCRGSHVLSVMVKTQCLVLLHLKHLPGHPVYRRKR	1454						
Db	425	CGSYRSVQBELTDMKQVFTNAEYVNCRGSHVLSVMVKTQCLVLLHLKHLPGHPVYRRKR	484						
QY	1455	KKFPDLAEDEGDSPEAVGQSRDEDRRSRE	1485						
Db	485	KKFPDLAEDEGDSPEAVGQSR--GRRQKK	513						
RESULT	26								
ID	ABR41346								
XX	AC	ABR41346							
XX	DT	02-JUN-2003 (first entry)							
XX	DE	Human DITHP transcription factor.							
XX	KW	Human; dithp; diagnostic and therapeutic polynucleotide; diagnosis;							
XX	KW	cancer; cell proliferative disorder; autoimmune disorder;							
XX	KW	inflammatory disorder; infection; hormonal disorder; metabolic disorder;							
XX	KW	neurological disorder; gastrointestinal disorder; transport disorder;							
XX	KW	connective tissue disorder; drug screening; proteome analysis;							
XX	KW	gene therapy; antisense therapy; genotyping; transgenic animal; knock in;							
XX	KW	disease model; toxicological testing; transcript imaging;							
XX	KW	transcription factor.							

XX OS Homo sapiens.
XX PN WO200297031-A2.
XX PD 05-DEC-2002.
XX PF 27-MAR-2002; 2002WO-US010056.
XX PR 28-MAR-2001; 2001US-0279619P.
XX PR 29-MAR-2001; 2001US-0280067P.
XX PR 29-MAR-2001; 2001US-0280068P.
XX PR 16-MAY-2001; 2001US-0291280P.
XX PR 17-MAY-2001; 2001US-0291829P.
XX PR 17-MAY-2001; 2001US-0291849P.
XX PR 19-JUN-2001; 2001US-0299428P.
XX PR 20-JUN-2001; 2001US-0299776P.
XX PR 20-JUN-2001; 2001US-0300001P.
XX (INCY-) INCYTE GENOMICS INC.
XX Daffo A, Jones AL, Tran AB, Dahl CR, Gietzen D, Chinn J;
XX Dufour GE, Hillman JL, Yu JY, Tuason O, Yap PE, Amshey SR;
XX Daughtery SC, Dam TC, Liu TF, Nguyen DA, Kleefeld Y, Gerstein EH;
XX Peralta CH, David MR, Lewis SA, Chen AJ, Panzer SR, Harris B;
XX Flores V, Marwaha R, Lo A, Lan RY, Urashka ME;
XX WPI; 2003-129518/12.
XX N-PSDB; ACC46286.
XX Novel human diagnostic and therapeutic polypeptide useful for identifying
XX test compound which specifically binds to a polypeptide encoded by human
XX diagnostic and therapeutic polynucleotide, and to induce antibodies.
XX Claim 27; SEQ ID NO 881; 591pp; English.
XX The invention relates to novel human diagnostic and therapeutic
XX polynucleotides designated dithp (ACC46080-ACC46749) and to their encoded
XX proteins (DITHP; ABR41136-ABR41812). The invention also relates to
XX polynucleotide sequences at least 90% identical to the dithp cDNA
XX sequences of the invention; recombinant vectors, host cells and
XX transgenic organisms comprising a dithp nucleic acid sequence; the
XX recombinant production of DITHP proteins; antibodies specific for DITHP
XX proteins; microarrays comprising dithp nucleic acid sequences; methods of
XX detecting dithp nucleotide and protein sequences; methods of screening
XX for compounds which specifically bind a DITHP protein; and methods of
XX assessing the toxicity of test compounds using a dithp hybridisation
XX probe. Dithp nucleic acid sequences and DITHP proteins may be used in the
XX diagnosis of a wide variety of conditions including cancer and other cell
XX proliferative disorders; autoimmune or inflammatory disorders; bacterial,
XX viral, fungal or parasitic infections; hormonal disorders; metabolic
XX disorders; neurological disorders; gastrointestinal disorders; transport
XX disorders; and connective tissue disorders. They may also be used to
XX screen for modulators of protein activity or gene expression. DITHP
XX proteins can additionally be used in analysis of the proteome of a tissue
XX or cell type and to induce antibodies. The dithp nucleic acids are
XX additionally useful in somatic or germline gene therapy of the disorders
XX mentioned above, as a source of antisense sequences, as a source of
XX probes and primers, in genotyping and identification of individuals, in
XX the generation of transgenic animal models of human disease or knock in
XX humanised animals, in toxicological testing, and in transcript imaging.
XX The present sequence represents a DITHP protein which has transcription
XX factor activity. Note: The sequence data for this patent did not form
XX part of the printed specification, but was obtained in electronic format
XX directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 209 AA;

Query Match 13.5%; Score 1078; DB 6; Length 209;
Best Local Similarity 96.7%; Pred. No. 3 4e-64;
Matches 204; Conservative 2; Mismatches 3; Indels 2; Gaps 1;
QY 1275 YEAVAGLRPRKTIIRGKHSVIPPAAARSGRRPGKPHSTRSQKAPPVDDAEVDELVLQ 1334

Db 1 YEVAGLRPRKTIKSGSVIPPAASGRPPGKSHSTRSPKAPPVDDAEVDELVLQT 60
QY 1335 KRSSRRQSLQKCEIILHKIVKRYFSPFPFVTRDEADYDVI THPMDPQTQVONKCS 1394
Db 61 KRSSRRQSLQKCEIILHKIVKRYFSPFPFVTRDEADYDVI THPMDPQTQVONKCS 120
QY 1395 CGSYRSVQEFLLDMKQVFNARVYNCGRSHVLSCHVKTQCLVLLHKLHGLPGHPYVRKR 1454
Db 121 CGSYRSVQEFLLDMKQVFNARVYNCGRSHVLSCHVKTQCLVLLHKLHGLPGHPYVRKR 180
QY 1455 KKFPPRLAEDGDSPEAVGQSRDEDRSRE 1485
Db 181 KKFPPRLAEDGDSPEAVGQSR--GRRQKK 209

RESULT 27
ID AA081169
ID AA081169 standard; protein; 1674 AA.
XX AA081169;
XX 05-MAY-1999 (first entry)
XX Human BAZ1-alpha protein.
XX Transcriptional regulator; transcription; BAZ1-alpha; bromodomain; BAZ;
KW atypical zinc finger; testis; human; tumour; BAZ1-beta; BAZ2-alpha; drug;
KW BAZ2-beta; treatment; cancer; proliferative disorder; screening.
XX Homo sapiens.
XX WO9847920-A1.
XX 29-OCT-1998.
XX 17-APR-1998; 98WO-JP001783.
XX 18-APR-1997; 97JP-00116570.
PR 24-OCT-1997; 97JP-00310027.
XX (CHUS) CHUGAI RES INST MOLECULAR MEDICINE INC.
XX Jones MH;
XX WPI; 1998-583603/49.
DR N-PSDB; AAV68385.

PT Transcriptional regulator gene family containing bromodomain - may be
PT expressed in testis tissue and is useful for treatment of cancer and
PT other proliferative disorders.
XX Claim 2; Page 46-58; 187pp; Japanese.
XX This sequence represents the human BAZ1-alpha protein, a member of a
CC family of transcriptional regulator genes containing a bromodomain (BAZ,
CC Bromodomain with Atypical Zinc Finger) which are expressed specifically
CC in testis tissue and also in certain tumour lines. Transgenic cells may
CC be used for the preparation of the BAZ1-alpha, BAZ1-beta, BAZ2-alpha and
CC BAZ2-beta proteins. These proteins can be used in the treatment of cancer
CC and other proliferative disorders, and in screening of compounds for
CC their binding ability to the expression products (e.g. for use as drugs
CC by modulation of transcriptional regulation)
XX Sequence 1674 AA;

Query Match 11.7%; Score 929.5; DB 2; Length 1674;
Best Local Similarity 23.1%; Pred. No. 68-53;
Matches 395; Conservative 264; Mismatches 529; Indels 525; Gaps 67;
QY 3 PLLGRKPF----PLVNPLPGEEPFFFTIPHTQAFRTREYEARELRYSERIWTCKSTGSS 58
Db 152 PLLHRKPFVQKPPADLRDEEVFY-CKVTNIEIFRHYDDFFERTILCNLSLVWSCAVTGRP 210

QY 59 QLTTHKEAWERQEAELKEBFPAYEKLVLVEMVH-----HNTASLEKLVDTAWLEIMTKY 114
Db 211 GLTYOEALLESKKARQNL-QSFP---EPLIIPVLYLTSLTHSRRLHEICDDIFAVVKDRI 266
QY 115 AVGECDEPVGKEMKLVKIVKIHPLKVDSEATEKKSDGACDPSDKENSSQIAQDHQ 174
Db 267 FVEETVEIRNNGARLOCTILEVLP-----PS-----HQNGFANGHV 303
QY 175 KK---ETVVKDEGRRESINDRARRSPKLPSTLKKGERKWAAPPKFLPHKVDV-----KLQ 227
Db 304 NSVDGETIISDSDSETQS-----CSFQNGKKDAIDPLL-FYKVVQPTKKEL 351
QY 228 NEDKIISNVPADSLIRTERPKNKEIVRYFIHNLALRAGTGENAPVWVEDELVKVYSLPS- 286
Db 352 HESAI---VKATQISRRKHLFSDKLKLFKQHC-----EPQEGVIK---IKASSLSTY 399
QY 287 ---KPSDFLLDPYKVMTLNPSTKRKNTGSPDRKPSKK---SKTDNSSLSPNPKLWC 338
Db 400 KIAEQDFSYFFDDPTTIFSPANRRRG-----RPPKRIHSQEDN-----440
QY 339 HVHLKLSLGSPLKYKNSKNSPPEHLEEMWKMSPNKLHTNFHIPKGGPPAKPGKHS 398
Db 441 -VANKQTLA-----SYRSKATKER-----458
QY 399 DKPLKAGRSKGIILNGQKSTGNSKSPKGLKTPKTKMKQMTLLDMAKGTQKTRAPRNSG 458
Db 459 DKLLK-----QEMKSLAFKAKLKR-----479
QY 459 GTPRTSSPKHLPAPAAHLIAYYKENDKDRDKSALSVCISKARTALLSSEDRARLPBEL 518
Db 480 -----EKADALEAKKKEKEDKKR-----BBL 502
QY 519 RSLVQRYELLEHKRWASMESEBQRYLKKRBELEKKLKEKAKERKEKEMLEKQK 578
Db 503 KKIVEEE-----RLKKKEKERLKYEREKEREKLEB--EKRYVEYLKQWSKPR 549
QY 579 RYEDDELTG-KNLPAFLVDTPEGLPNTLFGDVAMVVEFLSCYSGLLLPDAQYP--ITAV 635
Db 550 --EDMECDLDELPEPTPKT--RLPPEIFGDALMVLEFLNAPGELFQLOEFEPDGVTL 605
QY 636 SIMEAL-SADKGG-----FLVLRVLVILQTLQDRIAB-----DYGELGMKSEIPIJT 684
Db 606 VLEALVNDSEGPLCELLFFLTAIFQAIABE--EEVAKELTDATKCSLSLDD 663
QY 685 LHSVSELVRLCLRRSDVQSESGSTD-----DNKDSAAFEDNEVQDEFEKLETSF 737
Db 664 SCTLSEILRLHILASGADVTSANAKRYQKRGGFATDDACWELRLSNPSLVKLSSTSV 723
QY 738 FELTSEELQILTALCHRLIMTYSVD-----HMETEQQMSA 774
Db 724 YDLTPGEMKILHALCGKLLTLVSTRDFIEDYVDILROAKQEFRELKAEQHRKEEAAA 783
QY 775 ELWKERLAVLKEENDKKRAEKQKREMEAKN-----KENGKVENGLGKTRDKRIV 825
Db 784 RIRKKEKLEKEQKQKMEKEKLEKEDQORNSTADISIGEBEREDFTSIESKDTQK-- 841
QY 826 KPEQVDTAEADMSIAVSKRRLLAIQAK-----KEREIQE-----REMVKVLRQA-E 872
Db 842 ELDDQMFTEDEDDPGSHKRGK---RGKRGQNGKFEFTRQEQINCVTRELLTADBEALK 897
QY 873 EERIRKHAABAKAFQEGIAKAKLVMRTPITGTDNHNRYWLFSDVEVGLFTEK---GVV 929
Db 898 QEHQREKELELEK-IQSAIACNTIP-----PIGRDRMYRYWIF-PSIFGLFIEEDYSGIT 951
QY 930 HDSI-----DYRFNHCHCKDHTVSGDEYCPRSKKN-LGKQASMTQGTATEVAVETTT 983
Db 952 EDMLLPRSSQNNVQSDPQVS-----TKTGEPLMSESTNIDQG-FRDHSVQLPK 1002
QY 984 PKQGNLWFLCDSQKELDELLCNLHPQGIRESQLKERL--EK-----RYQDIHSHI 1033
Db 1003 PVHKPNRCWFYSSCBQLDQLEALNSRGHRESALKETLLQEKSRICACQLARPSE--EKPH 1060

Db	606	VLEALVGNDS	EGPLCELLFFLTAIFQAI	AEE--EEVAKEQLTADT	RGCSLKUSDLD	663										
Qy	685	LHYSSELVRL	CLRRSDVQES	EGSDTD-----DNK	SDAAAFNEVODEFLEKLETSEF	737										
Db	664	CTULSELRL	HLHILASGADVT	SANAKYRKRG	GFGFDATDDACMELRLSNP	SLVLSKSTSV	723									
Qy	738	FELTSEKQL	ILTALCHRI	LMTYSVOD-----	HMETROQMSA	774										
Db	724	YDLTPG	EKMILHALCGK	LLTVSTRDFIEDYVD	ILRQAQEPRELKAEQHR	KEREBEAAA	783									
Qy	775	ELWTERLAV	LKEENDKRAE	KQKKEWAKY-----	KENGKVENG	LKTKDTRKKRIV	825									
Db	784	RIRKREKLE	KEQKQKKEQ	KEKLKDEQRNSTADIS	IGEEERBDFPTSIESK	DTQEQ--	841									
Qy	826	KFEPQVTE	AEADMISAVK	SRELLAIQAK-----	KEREIOE-----	REMKVKLEROA	E	872								
Db	842	ELDQDMTE	DEDDPGSKRGR	----RGRKONG	FEKFTREQINCV	TRELLTAD	EEBALK	897								
Qy	873	EERIRKHAA	EAKAFQEGIA	KAKLVMRBTPIGT	DNRHNYLWFSDE	VPGLFIEK--	--GWV	929								
Db	898	QEHQKEKEL	EK-IQSAI	ACTNIF----PLGRD	RMVRYWIP-PSI	PGLFIEDY	SGLT	951								
Qy	930	HDSI-----	DYRPNHCKD	HTVSGDEYCP	RSKAN-LGNKAS	NTOHGTATE	VAVETTT	983								
Db	952	EDMLLP	PPSSFO	NNVQSDPQVS-----	TKTGEPL	MESEST	SNIDQG-PROHS	VOLPK	1002							
Qy	984	PKOGONL	WFLCDSOKEL	DELINCLHP	OGIRESOLKERL--	EK-----	RYODI	IHSIH	1033							
Db	1003	PVHKPNK	RCYSSCEQ	LDLLEALNS	GHRESALKETL	LOKSRI	CAQLARFSE--	EKFH	1060							
Qy	1034	LARKPN	GLK-----	SCDGNQ-----	ELLNFLR	SLDIEVAT	RLOKGLGYVEET-	1077								
Db	1061	FSDKPD	QSDPTYS	GRSSNAYDPS	QCAEQLERL	DFLLDIED	RIYQOTL	GAIK	1120							
Qy	1078	-----SFEA	---RVISLEK	-----KDFGEC	VIQASVIXK	FLOGF	WAPKQRR	1120								
Db	1121	RHIWRS	ALSGRYELL	SEENKENGII	KTVNEDVE	EMEIDEQ	TKVIVK--	DRLLG	IKTETP	1178						
Qy	1121	KLOQSD	AKTEVE	D-----	EKKWVEE	AKVAS-----	ALEKWKTAI	1156								
Db	1179	STVSTN	ASTPQSV	SVVHYLAMAL	FQTEQ	GIERRFL	KAPLDAS	DGSR	YKTVLDR	WRESL	1238					
Qy	1157	REAGTFR	SMVLLQ	MDACIKW	MSAENARCKV	CPKGEDD	KLLJL	CDCKNA	FHLC	RLP	1216					
Db	1239	LSSASL	SQVFLH	LTLDRS	VIWSKSI	ILNARCKI	CRKGD	AEANVVL	CDGCD	RHHITYC	VRP	1298				
Qy	1217	ALYEV	PDGWC	CPACQ	PATARRN	SRGRNYTE	ESASE	SD	DEDEE	-----	EEEEEEE	1272				
Db	1299	KLKTV	PGDW	FCPC	RCRQR	RLRSLFR	PSLESD	EDVDS	MGDEDD	EV	DGDEEG	QOSE	1358			
Qy	1273	EDYEV-----	AGRLR	PRKTI-----	RGKHS	VIPPAARS	GRKPG	KPKHSTR	1314							
Db	1359	EYEV	EODEDD	SQEE	EVSLPKRGP	PQVRLPVK	TRGLK	SSPS	RGQOQ	EPGRY	PSR	SOQ	1418			
Qy	1315	SQPK-----	APPVD-----	DAE	DEL-----	1330										
Db	1419	STPKTT	VSSKTR	SLRKINS	APPTETK	SLRSTR	SHSGPL	QADVF	VELL	SPR	KRRG	1478				
Qy	1331	-----	VLQTK	ESSR	ROSL-----	1343										
Db	1479	RKSAN	TNPENS	PNPN	FRVJATKS	ESQSR	SVNTASK	LSLO	ESKRR	CRK	QSP	PEPS	SVT	1538		
Qy	1344	-----	ELQKCE	BITLHKI	VKRYFS	WPPEV	PTVTR	DEAD	EYDVITH	PM	DF	TVQ	1390			
Db	1539	LGRSS	CRQGV	HEL	SAFELQ	VLVLR	HRD	SSWPL	KLVSK	IOVP	PDYD	YDI	IKK	IALNI	IR	1598
Qy	1391	NKSCG	GYRS	VQ	EFLTM	QVNTMA	VYNC	RG	1423							
Db	1599	EKNKCY	KLASE	FID	DIEL	MFNS	CN	FY	NP	RP	NT	1631				

ADP12578	ADP12578 standard; protein; 1674 AA.
ID	ADP12578;
AC	ADP12578;
XX	
XX	
DT	12-AUG-2004 (first entry)
XX	
DE	Protein encoded by mRNA of the invention #188.
XX	
KW	transplant rejection; immune system; rheumatoid arthritis; lupus;
KW	inflammatory bowel disease; multiple sclerosis; HIV; AIDS.
XX	
OS	Homo sapiens.
XX	
FN	WO2004042346-A2.
XX	
PD	21-MAY-2004.
XX	
PF	24-APR-2003; 2003WO-US012946.
XX	
XX	24-APR-2002; 2002US-00131831.
PR	20-DEC-2002; 2002US-00325899.
XX	
PA	(EXPR-) EXPRESSION DIAGNOSTICS INC.
XX	
FI	Wohlgemuth J, Fry K, Woodward R, Ly N, Prentice J, Morris M;
FI	Rosenberg S;
XX	
DR	WPI; 2004-400724/37.
XX	
PT	Diagnosing or monitoring transplant rejection, e.g. heart, kidney, liver,
PT	pancreas, pancreatic islet, lung, bone marrow or stem cell transplant
PT	rejection, in an individual, comprises detecting the expression level of
PT	the genes.
XX	
PS	Claim 65; SEQ ID NO 2587; 1762pp; English.
XX	
CC	The present invention relates to diagnosing or monitoring transplant
CC	rejection, e.g. cardiac or kidney transplant rejection, in an individual
CC	comprises detecting the expression level of one or more genes. The
CC	methods, system and kits are useful in diagnosing or monitoring
CC	transplant rejection, e.g. heart, kidney, liver, pancreas, pancreatic
CC	islet, lung, bone marrow or stem cell transplant rejection,
CC	xenotransplant rejection or mechanical organ replacement rejection, in an
CC	individual. The method is also useful in assessing the immune status of
CC	an individual. The methods are also useful in diagnosing and monitoring
CC	diseases that involve the immune system, e.g. rheumatoid arthritis,
CC	lupus, inflammatory bowel diseases, multiple sclerosis, HIV/AIDS or
CC	viral, bacterial or fungal infection. The present sequence represents a
CC	protein that is encoded by the mRNA of the invention.
XX	
SQ	Sequence 1674 AA;
	Query Match 11.7%; Score 929.5; DB 8; Length 1674;
	Best Local Similarity 23.1%; Pred. No. 6e-53;
	Matches 395; Conservative 264; Mismatches 529; Indels 525; Gaps 67;
Qy	3 PLLGRKEP----PLVNPPLGPEEPFTTPIHTQEAFTREEYEYEARLRYSEIRWTCKSTGSS 58
	152 PLLHRKEFPVQKPPADLRPDDEVFY-CKVTNEIFRHYDDPFERTILCNSLVWSCAVTRGP 210
Db	
Qy	59 QLTHKEAWEEQVAEILLKEEFPANWEKIVLEWVH----HNTASLEKLVDTAWLEIMTKY 114
	211 GLTYQEALESEKKARQNL-OSFP---EPLIIPVLVLTSLTHRSRLHEICDDIFAYVXDY 266
Db	
Qy	115 AVGECCDFEVGKEKMLVKVIVKIHPLKVDDEATEKSGDCACDPSKDKNSQIAODHQ 174
	267 FVEETVEIRNGARLQCTTILEVP-----PS-----HQNGFANGHV 303
Db	
Qy	175 KK----ETVVKEDGREGRESINDRARRSPRKPLTSLKKGERKWAQPKFLPHKYDV-----KLQ 227
Db	304 NSVDGETIITISDSDSSTQS-----CSFQNGKKKDAIDPLL-FYKVPQPTKKEL 351

QY 228 NEDKIISNVADSLIRTERPPNKEIVRYFIRNRLRAGTGENAPWVVEDELVKYSLPS- 286
DB HESAI---VKATQISRRKHLFRGDKLFLKQHC-----EPQEGVIK---IKASSLSY 399
QY 287 -----KESDPLDLPYKMTLNPSTKXKNTGSDRPSKK---SKTDSLSLSPINPKLWC 338
DB 400 KIAEQDSYFFPDPTFIPSPANRRG-----RPPKRIHISQBDN----- 440
QY 339 VHYLKKSLSGSLPKVKNKSNKSPBEHLEEMKMMSPNKLHFNHFKKGGPPAKKPGKHS 398
DB 441 -VANKOTLA-----SVRSKATKER----- 458
QY 399 DKPLKAKGRSKGILNGQKSTGNSKSPKGLKTPKTKYMKQMTLLDMAKGTQKTRAPRNSG 458
DB 459 DKLLK-----QBEKSLAFKAKLKR----- 479
QY 459 GTPRTSSKPKHKLPPAALHLIAVYKENKDREKRSALSVCISKTARLLSSEDRARLPEEL 518
DB 480 -----EKADALEAKKKEDEKKR-----BEL 502
QY 519 RSLVQRYELLEHHKRWASMSBEORKEVLLKKREELKKLKEKAKERKEKEMLEKQK 578
DB 503 KKIVEE-----RLKXKEKERLKVREKEREKLE--EKRYKVEVLKQWSKPR 549
QY 579 RYEDQELTG-KNLPAFRLVDTPEGLPNTLFGDVAMVVFELSCYGLLLPDAQYP--ITAV 635
DB 550 --EDMECODLKELPPTPKT--RLPPEIFGDALVLEFLNAFGLFDLQDFPDGVTLE 605
QY 636 SLMEAL-SADKGG-----FLYNRVLLTQTLQDEIAE-----DYGELGNKLSIEPLT 684
DB 606 VLEEALVGNDSGPELCELLFFFLTAFQAIABE--EEVAKBQLTADTKGCSLKSLLD 663
QY 685 LHSVSELVRLCLRRSDVQESGSDT-----DNKDSAAFEDNEVDQEFLEKLETSEF 737
DB 664 SCTLSEILRLHLASGADVTSANAKRYQKRGFDATDDACWELRLSNPVLKLSSTSV 723
QY 738 FELTSEKQLQILTALCHRLMTYSVD-----HMETRQMSA 774
DB 724 YDLTPEKMKILHALCGKLLTLVSTRDFTYVDILRQAKQEFRELKABQHRKEREAAA 783
QY 775 ELWKEPLAVLKEENDKRAEKQKREMEAKN-----KENGKVENGLGKTDKRIIV 825
DB 784 RIRKKEEKLKEQEKQKMEKQKLEDEQNRSTADISIGEEERDFTSIESKDEQK-- 841
QY 836 KFEPOVDTEAEDMISAVKSRRLALQAK-----KERETQE-----REMKVKLERQA-E 872
DB 842 ELDDQMFTEDEDDPGSHKGR-----RGKGQNGFKEFTRQEQINCVTRELTLTADSEALK 897
QY 873 EERIRKHAABAKAFQEGIAKAKLVMRRTPIGTDNRNRYWLFSDVFPGLFIEK---GWV 929
DB 898 QEHQKEKELLEK-IQSAIACNIF-----PLGRDRMYRYWIF-PSIPLGFIEEDYSGLT 951
QY 930 HDSI-----DYRFNHCKDHTVSGDEDCYCPRSKAN-LQKNASMTQGTATEVAVETT 983
DB 952 EMLLPSPFSFQNNVQSDPQVS-----TKTGEPLMSESTSIDQG-PRDHSVOLPLK 1002
QY 984 PRQGNLFLCDSOKELDELNCLHPQGIRESOLKRL--EK-----RYODIHSIH 1033
DB 1003 PVHKENRWCFYSQOLQOLIBALMSRGHRESALKETLLQEKSRICAQLARFSE--EKPH 1060
QY 1034 LARKPNLGLK-----SCDGNQ-----ELLNPLRSDLIEVATRLQKGLGYVEET- 1077
DB 1061 FSDKQPDSPKTPYSRGRSSNAYDPSQMAEKQLELRDLRDLADIEDRIYQGTGAIKVTD 1120
QY 1078 -----SEFEA-----RVISLEKL-----KDFGECVIALQASVKKFLQGFMAPKQKR 1120
DB 1121 RHIMRSALSGRYELLSEENKENGIIKTVDNEVBEIMEIDQTKVIVK--DRLLGIKTETP 1178
QY 1121 KQSDSAKTEVDE-----EKKWVEEAKVAS-----ALEKWKTAI 1156
DB 1179 STVNASTNQSVSSVWVYALMALFQIBQIEERFLKAPLADSDGRSYKTVLDWRRESL 1238
QY 1157 REAQTFSRMVLLGLMDACIKWMSAENARCKVCPKKGEDDKLILCDECNKAFHLFCLRP 1216

DB 1239 LSSASLSQVFLHLSTLDRSVIWSKILNARCKICRKGDAENMVLCDGDRGHHTYCVRP 1298
QY 1217 ALYVEPDGEWQCPACOPATARNRGRNVYTBESSASDESDDEDEEE-----EEEEEEE 1272
DB 1299 KLVTVPEGDMFCPECRKQRCRRLFRORPSLESDDEVEDSGGSDDEVDGDEEGQSEE 1358
QY 1273 EDYEV-----AGLRRLRPKTI-----RGHSVIPPAARSGRRPGKPKHSTRR 1314
DB 1359 EEEYEQDEDDQSEEEVSLPKRGRPOVRLPVKTRGKLSLSSFSRGGQQQEPGRYPSRSQ 1418
QY 1315 SQPK-----APPVD-----DAEVEL----- 1330
DB 1419 STPKTTVSSKTRSLRKINSAPPTETKSLRIASRSTRHSHGLOADVFFVLLSPRRKRG 1478
QY 1331 -----VLQTKRSRRQSL----- 1343
DB 1479 RKSANNTPENSFPNFPNFRVIATKSEQSRSVNIASKLSQESKRCRKRSPPSPVT 1538
QY 1344 -----ELQKCEILHKIVKYRFSFPREPVTTRDEADYDYDVITHPMDFQTVQ 1390
DB 1539 LGRSSRGGQGVHELSAFEQLVVELVRHDDSWPFLKLVSKIQVPDYDIKKPIALNIIR 1598
QY 1391 NKSCGYSYRVOEFLTDMKQVFTNAEVNCRGS 1423
DB 1599 EKVNKCEYKLASEFIDDIIDIELMFNSCFEYNPRNT 1631

RESULT 30

AA07734

ID AA07734 standard; protein; 1540 AA.

AC AA07734;

DT 02-JUL-1999 (first entry)

DE Human hTLP2 protein.

KW Telomerase protein; hTLP2; human; biocontrolling mechanism; cell growth;
cell aging.

OS Homo sapiens.

PN JP11089579-A.

PD 06-APR-1999.

PF 24-SEP-1997; 97JP-00258450.

PR 24-SEP-1997; 97JP-00258450.

PA (MITU) MITSUBISHI CHEM CORP.

XX WPI; 1999-290828/25.

XX N-PSDB; AAX37304.

PT A higher animal telomerase protein and a gene coding it.

PS Example 1; Page 12-18; 22pp; Japanese.

CC This invention describes a novel human telomerase protein, hTLP2. This
protein and the gene coding it are useful for the elucidation of
biocontrolling mechanisms such as cell growth and cell aging

SQ Sequence 1540 AA;

Query Match 11.6%; Score 924.5; DB 2; Length 1540;

Best Local Similarity 23.0%; Pred. No. 1.2e-52;

Matches 397; Conservative 264; Mismatches 532; Indels 531; Gaps 67;

QY 3 PLLGRKPF-----PLVNPLPGSEPPPTIPIHTQEAFTREYEARELRYSERIWTCKSTGSS 58

DB 2 PLLHKKPFVRQKPPADLRPDEEVFY-CKVTNEIFRHYDDFFERTILCSLVMSCAVTGRP 60

QY 59 QLTKEAMEEEOVAELLKEBPANYEKIVLEWH-----HNTASLEKLVDTAWLEIMTKY 114
Db 61 GLTYOEALLESSEKARQNL-QSFP-----ELIIPVLYLTSLTHRSRUIHEICDDIFAVKDRY 116
QY 115 AVGECDFEVGKEMKLVKVIKIHPLEKVDEATEKSGDAGCDSPSKENSSQIAQDHQ 174
Db 117 FVEETVEVRNNGARLQCLILEVP-----PS-----HQGFANGHV 153
QY 175 KK-----ETVVKDEGRRESINRARRSKPLPTSLKKGRKWKAPPKFLPHKYDV-----KLQ 227
Db 154 NSVDGETIIISDSDSETQS-----CSFQNGKKDAIDPLL-FYKVKQPTKKEL 201
QY 228 NEDKIISNVPADSLIRTERPPNKEIVRYFIRHNALRAGTGENAPVWVEDELVKYSLPS- 286
Db 202 HESAI-----VKATQISRRKHLFSRDLKLFLKQHC-----EPQDGVIK-----IKASSJSTY 249
QY 287 -----KFGDFLLDPKYMTLPSTPKRNTGSPDRKPSKK-----SKTDSNLSPLNPKLWC 338
Db 250 KIAEQDFSYFFPDPTTIFSPANRRG-----RPPKRIHSQEDN----- 290
QY 339 HVHLKKSLSGSLPKVKNKSKSPBEHLEEMKMMKSPNKLHTNFHIPKKGPPAKPKGRHS 398
Db 291 -VANKOTLA-----SYRSKATKER----- 308
QY 399 DKPLKAKGRSGILNGQSTGNSKSPKGLKTPKTKMKQMTLLDWAQTKQMTAPRNSG 458
Db 309 DKLLK-----QSEMSLAPEKAKLKR----- 329
QY 459 GTPRTSKPHKHLPPAALHLIAYKENKDRDKRSALS CVISKTA RLSSSDRAKLPEEL 518
Db 330 -----EKADALEAKKKEDEKKEK-----EEL 352
QY 519 RSLVOKRYELLEHKRWASMEEOKEYLKKKBEELKKLKEKAKERKEKEMLEKQK 578
Db 353 KKIIVEE-----RLKKKEKERLKVEREKEREKLE- -EKRYVEYLKQWSKPR 399
QY 579 RYEDQELTG-KNLPAPRLVDPTEGIPNTLFGDVANVVEFLSCYSGLLLPDAQYP--ITAV 635
Db 400 --EDMECDLKLPEFTPVKT--RUPPEIFGDALMVLEFNAFGELFDLQDEFDPGVTL 455
QY 636 SIMEAL-SADKGG-----FLYNLRVLVILLQTLLODRIAE- - 670
Db 456 VLEALVGNDEGPICELLFFLTAIFQAIABE--EEVAKELTADTKDLTEALDEDA 513
QY 671 -----YGEL--GMKLSIPIPLTHSVSELVRLCLRRSDVQSESGSDTD- 711
Db 514 DPTKSALS AVASALAAAWPQLHQGCSLSLDLSCTLSLRLHLILASGADVTSANAKRY 573
QY 712 -----DNKDSAAFEDNEQVDFLEKLETSFFELTSEEKQILTLCHRLIMTYSVOD- 764
Db 574 QKRGFDATDDACMBELRLSNPSLVKLSSTSVYDITPGEKMLHALCGKLLTLVSTRDF 633
QY 765 -----HMETROOMGAELWKEKELAVLKEENDKKRAEKQKKEKME 802
Db 634 IEDYVDILROAQEPRELKACHRKEREAAARIKRKEELKEQEQKQKQKLEKDE 693
QY 803 AKNKENGKVENGLGKTRDKRIVKPEPQVDTA-----EDMI-----SAVKGRRLAI 850
Db 694 ---QRNSTADISIGEEREDFDTIESK-DTEQKELDQDMVTEDEDDPGSHKGRGRKRG 749
QY 851 Q-AKKERIOEREMKVKLE--RQAEERIRKHAAAEKAFQEGIAKAKLVNRRPTIGTDR 907
Db 750 QNGFKEFTROQINCVTREPLTADSEALKQEHQKREKELLEKIQSAIACTNIPPLGRDR 809
QY 908 NNNRWLFSDEVPGLFIEK--GWVHDSI-----DYRNNHCKOHTVSGDEDCYCRSKA 959
Db 810 MYRRWIF-PSIPGLFIEDYSGLTEDMLLPRPSSFQNNVQSDPQVS-----TKTG 860
QY 960 N-LGKNASNTQHGATVAVETTPKQGNLWFLCDSQKLELNLCHLPQGIRESOLK 1018
Db 861 EPLMSESTSNIDQG-PRDHSVQLPKPVHKPNRWCFYSSCEQLDQLIALNSGRHRESALK 919

QY 1019 ERL--EK-----RYODIIHSIHLARKPNLGLK-----SCDGN-----ELLN 1053
Db 920 ETLQEKSRICAQIARSE--EKHFSDKPODPKPYYSRGRSSNAYDPSQWCAEKQLEL 977
QY 1054 FLRSLIEAVTRLOKGLGYVEET-----SEFEA--RVISLEKL-----KDFGE 1095
Db 978 RLRFLLDIEDRIYQGTGALKVTDRIHWSALESGRYVELLSEENKENGIIKTVNEDVEE 1037
QY 1096 CVIALQASVIKFLQGFMAPKQKRKLQSEDSAKTEEVDE-----EKMM 1139
Db 1038 MEIDEQTKVIVK--DRLLGIKTTETPSTVSTNASTPQSVSVVHYLAMALFQIEOGIERRF 1095
QY 1140 VEEAKVAS-----ALEKWKTAIRBAQTSRHHVLLGLMDACITKWDMSAENARCKVCP 1191
Db 1096 LKAPLDASDSGRSVKTVLDRWRBSLSSASLSQVFLHLSTLDRSVIWSKSLNARCKICR 1155
QY 1192 KKGEDDLILCDECNKAFHLFCLRPALYVEPDGEWQCQCPACQATARRNSRGRNTEESAS 1251
Db 1156 KKGDAENWVLCDCDGRGHHTYCVRPKLTVPEDGWFCECPKQKRSRRLSSRQPSLESD 1215
QY 1252 EDSDDSDSEEEE-----EEEEEEEDYEVA-----GLRLRPKTIIRGHSHVIPPAAARSGR 1303
Db 1216 EDVEDSGEGEDVDGDEEGQSEEEVEVEQDEDDSL-EVKT-RGKLSSFSRSGQQQ 1273
QY 1304 RPKKPHSTRESQPK-----APVD-----DAEV 1327
Db 1274 EPGYPSRSQOSTPTTVSSKTGRSLRKINSAPTETKSLRIASRSTRHSHGFLQADV 1333
QY 1328 DEL-----VLQTKRSTRRQSL----- 1343
Db 1334 ELLSPRRKRRGRKSANNTPENSFPNFRVIATKSSQSRSVNIASKLSLQSESKRRCR 1393
QY 1344 -----ELOCSEILHKIVKYRFSWPPREPVTDEADYDV 1379
Db 1394 KRQSPSPVTLGRSSRGQGVHLSAFEQLVVELVRHDDSWPFLKLVSQIQVDDYDI 1453
QY 1380 ITHPMDFTQNKCSGYSRSVQSFELTDMKOVFTNAEYVNCGRS 1423
Db 1454 IKKPIALNIIRKYNKCEYKLASEFIDIELMFNSCFEYNPNT 1497
RESULT 31
AAU16181
ID AAU16181 standard; protein; 141 AA.
XX
AC AAU16181;
XX
DT 07-NOV-2001 (first entry)
XX
DE Human novel secreted protein, Seq ID 1134.
XX
KW Human; immunosuppressive; antiarthritic; antirheumatic; cytostatic;
KW cardiant; vasotropic; cerebroprotective; nootropic; neuroprotective;
KW antibacterial; virucide; fungicide; opthalmological; vulnerary;
KW secreted protein; rheumatoid arthritis; hyperproliferative disorder;
KW cardiovascular disorder; cardiac arrest; cerebrovascular disorder;
KW cerebral ischaemia; angiogenesis; nervous system disorder;
KW Alzheimer's disease; infection; ocular disorder; corneal infection;
KW wound healing; epithelial cell proliferation; skin ageing; food additive;
KW preservative; antiproliferative.
XX
OS Homo sapiens.
XX
PN WO200155322-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US0011341.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.

PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225457P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231142P.
PR 08-SEP-2000; 2000US-0231143P.
PR 08-SEP-2000; 2000US-0231144P.
PR 08-SEP-2000; 2000US-0231143P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234977P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.

PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 17-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0251989P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.

XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-488783/53.
XX N-PSDB; AAS26168.

PT New nucleic acid molecules encoding 461 human secreted proteins for
PT diagnosing, preventing, treating or ameliorating medical conditions and
PT used as food additives or preservatives.
XX Claim 11; SEQ ID NO 1134; 980pp; English.
XX

X

CC The invention relates to isolated nucleic acid molecules and their
 CC encoded secreted proteins. The nucleic acids and proteins are used to
 CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,
 CC rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used
 CC in diagnosing a pathological condition or susceptibility to a
 CC pathologic condition. Antibodies to the proteins can also be used in
 CC alleviating symptoms associated with the disorders and in diagnostic
 CC immunoassays e.g. radioimmunoassays or enzyme linked immunosorbant assays
 CC (ELISA). Disorders which are diagnosed or treated include autoimmune
 CC diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g.
 CC neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac
 CC arrest, cerebrovascular disorders e.g. cerebral ischaemia, angiogenesis,
 CC nervous system disorders e.g. Alzheimer's disease, infections caused by
 CC bacteria, viruses and fungi and ocular disorders e.g. corneal infection,
 CC and many other disorders listed in the specification. The polypeptides
 CC can also be used to aid wound healing and epithelial cell proliferation,
 CC to prevent skin aging due to sunburn, to maintain organs before
 CC transplantation, for supporting cell culture of primary tissues, to
 CC regenerate tissues and in chemotaxis. The polypeptides can also be used
 CC as a food additive or preservative to increase or decrease storage
 CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,
 CC minerals, cofactors and other nutritional components. The present
 CC sequence represents a novel secreted protein of the invention. Note: The
 CC sequence data for this patent did not form part of the printed

Query Match 8.9%; Score 712; DB 4; Length 141;
 Best Local Similarity 99.3%; Pred. No. 7.9e-40;
 Matches 140; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 464 SSKPHKLPAAHLIAIYKKNDRKRSALSCVISTARLLSSDRARLPPEELSLVQ 523
 Db 1 SSKPHKLPAAHLIAIYKKNDRKRSALSCVISTARLLSSDRARLPPEELSLVQ 60
 QY 524 KYELLEHKRWASMSSEQRKYLKKRBEELKKUKKAKERREKEMLERLEKQRYEDQ 583
 Db 61 KYELLEHKRWASMSSEQRKYLKKRBEELKKUKKAKERREKEMLERLEKQRYEDQ 120
 QY 584 ELTGKNLPFAFLVDTPEGLPN 604
 Db 121 ELTGKNLPFAFLVDTPEGLPN 141

RESULT 32
 ABUS5250
 ID ABUS5250 standard; protein; 141 AA.
 AC ABUS5250;
 DT 18-MAR-2003 (first entry)
 XX Human novel polypeptide #337.
 KW Human; neural disorder; immune system disorder; renal disorder;
 KW muscular disorder; respiratory disease; reproductive disorder;
 KW gastrointestinal disorder; pulmonary disorder; cardiovascular disorder;
 KW hyperproliferative disorder; inflammatory disease; allergic reaction;
 KW blood related disorder; cancer; immunosuppressive; antiinflammatory;
 KW cardiovascular; nephrotropic; cytostatic; antiallergic; thrombolytic;
 KW haemostatic; antiarteriosclerotic.
 XX
 OS Homo sapiens.
 XX
 XX US2002132753-A1.
 PN 19-SEP-2002.
 PD
 XX 17-JAN-2001; 2001US-00764864.
 PF
 XX 31-JAN-2000; 2000US-0179065P.
 PR 04-FEB-2000; 2000US-0180628P.
 PR 28-JUN-2000; 2000US-02114886P.
 PR 07-JUL-2000; 2000US-0216647P.
 PR 07-JUL-2000; 2000US-0216880P.

PR 11-JUL-2000; 2000US-0217487P.
 PR 11-JUL-2000; 2000US-0217496P.
 PR 14-JUL-2000; 2000US-0218290P.
 PR 26-JUL-2000; 2000US-0220963P.
 PR 26-JUL-2000; 2000US-0220964P.
 PR 14-AUG-2000; 2000US-0224518P.
 PR 14-AUG-2000; 2000US-0224519P.
 PR 14-AUG-2000; 2000US-0225267P.
 PR 14-AUG-2000; 2000US-0225268P.
 PR 14-AUG-2000; 2000US-0225270P.
 PR 14-AUG-2000; 2000US-0225447P.
 PR 14-AUG-2000; 2000US-0225757P.
 PR 22-AUG-2000; 2000US-0225758P.
 PR 30-AUG-2000; 2000US-0226868P.
 PR 01-SEP-2000; 2000US-0228924P.
 PR 01-SEP-2000; 2000US-0229287P.
 PR 01-SEP-2000; 2000US-0229343P.
 PR 01-SEP-2000; 2000US-0229344P.
 PR 01-SEP-2000; 2000US-0229345P.
 PR 05-SEP-2000; 2000US-0229509P.
 PR 08-SEP-2000; 2000US-0229513P.
 PR 08-SEP-2000; 2000US-0231413P.
 PR 21-SEP-2000; 2000US-0234223P.
 PR 21-SEP-2000; 2000US-0234274P.
 PR 25-SEP-2000; 2000US-0234997P.
 PR 27-SEP-2000; 2000US-0235834P.
 PR 29-SEP-2000; 2000US-0236327P.
 PR 29-SEP-2000; 2000US-0236367P.
 PR 29-SEP-2000; 2000US-0236368P.
 PR 29-SEP-2000; 2000US-0236369P.
 PR 29-SEP-2000; 2000US-0236370P.
 PR 02-OCT-2000; 2000US-0236802P.
 PR 02-OCT-2000; 2000US-0237037P.
 PR 02-OCT-2000; 2000US-0237038P.
 PR 02-OCT-2000; 2000US-0237039P.
 PR 02-OCT-2000; 2000US-0237040P.
 PR 13-OCT-2000; 2000US-0239935P.
 PR 20-OCT-2000; 2000US-0240960P.
 PR 20-OCT-2000; 2000US-0241785P.
 PR 20-OCT-2000; 2000US-0241809P.
 PR 01-NOV-2000; 2000US-0244617P.
 PR 17-NOV-2000; 2000US-0249299P.
 PR 08-DEC-2000; 2000US-0251856P.
 PR 08-DEC-2000; 2000US-0251868P.
 PR 08-DEC-2000; 2000US-0251869P.
 XX (ROSE/) ROSEN C A.
 PA (RUBE/) RUBEN S M.
 PA (BARA/) BARASH S C.
 XX Rosen CA, Ruben SM, Barash SC;
 XX WPI; 2003-147444/14.
 DR N-PSDB; ABX73509.
 XX New polypeptides and nucleic acids, useful in gene therapy for treating,
 PT inhibiting or preventing e.g. neural, immune system, muscular,
 PT respiratory, reproductive, gastrointestinal, pulmonary, cardiovascular or
 PT renal disorders.
 XX
 PS Claim 11; SEQ ID NO 1134; 402pp; English.
 XX
 CC The invention relates to human novel polypeptides and their associated
 CC polynucleotides. The polypeptides and polynucleotides are useful in gene
 CC therapy for treating, inhibiting or preventing neural disorders, immune
 CC system disorders (e.g. systemic lupus erythematosus, rheumatoid arthritis
 CC and multiple sclerosis), muscular disorders, respiratory diseases (e.g.
 CC nasal vestibulitis, nasal polyps and sinusitis), reproductive disorders,
 CC gastrointestinal disorders, pulmonary disorders, cardiovascular disorders
 CC (e.g. congenital heart defects, Ebstein's anomaly and hypoplastic left
 CC heart syndrome), renal disorders (e.g. acute kidney failure and end-stage
 CC renal disease), hyperproliferative disorders (e.g. Hodgkin's disease and
 CC leukaemia), inflammatory diseases (e.g. septic shock, bursitis and

CC appendicitis), allergic reactions and conditions (e.g. asthma), blood
CC related disorders (e.g. thrombosis, atherosclerosis and myocardial
CC infarction) and cancerous diseases. Sequences ABUS4914-ABUS5699 and
CC ABUS5748 represent human novel polypeptides of the invention
XX
SQ Sequence 141 AA;

Query Match 8.9%; Score 712; DB 6; Length 141;
Best Local Similarity 99.3%; Pred. No. 7.9e-40;
Matches 140; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 464 SSKPHKLPALHIIAYYKENDKEDKESALSCVISTARLLSSEDRARLPEELSLVQ 523
D 1 SSKPHKLPALHIIAYYKENDKEDKESALSCVISTARLLSSEDRARLPEELSLVQ 60

QY 524 KYELLEHKRWASMEQEKYLLKKREELKKLKKERREKEMLERLEKQRYEDQ 583
D 61 KYELLEHKRWASMEQEKYLLKKREELKKLKKERREKEMLERLEKQRYEDQ 120

QY 584 ELTGKLPAPRLVDTPEGLPN 604
D 121 ELTGKLPAPRLVDTPEGLPN 141

RESULT 33
ABB58706
ID ABB58706 standard; protein; 1476 AA.
XX
AC ABB58706;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 2910.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US009231.
XX
PR 23-MAR-2000; 2000US-0191637P.
PR 11-JUL-2000; 2000US-00614150.
XX
PA (PEKE) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
XX
XX WPI: 2001-656860/75.
DR N-PSDB; ABL02809.
XX
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions.
XX
PS Disclosure; SEQ ID NO 2910; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from wipo.int/pub/published_pct_sequences
XX
SQ Sequence 1476 AA;

Query Match 8.3%; Score 664; DB 4; Length 1476;
Best Local Similarity 18.3%; Pred. No. 3.5e-35;
Matches 330; Conservative 255; Mismatches 502; Indels 714; Gaps 56;

QY 3 PLLGRKPPPLVNPDPGEPPF-----FTIPHTQEAFFRTREYEARELRYSEIRWTCKSTG 56
D 2 PICKREGFDLQKQEGKNETFHDNDQVCCYITKRIFRDYEHYFRHVMVINSTVMOCEATG 61

QY 57 SSQLTAKAWBEEQVABLLKEEPPAWYEKLVLEVMVHNTASLEKLVDTAWLEIWTK-YA 115
D 62 KENLYTEAVKSER-AARKQMEQFQOSLRAPVLLVVEHAQSAVNTLNMIVAKFKRYP 120

QY 116 VGEEDCFVGKEMKLVKVIKIHIPLEKVDDEATEKKSDGACDPSDSSQIAQDHQK 175
D 121 IGEESVQAKXNATVTVLGVL-----DKNMPPLNGIYED 156

QY 176 KETVKEDEGRRESINDRARRSPKLTSLKKGKRWAPPKPLPHKYDVKLQNEKDIISN 235
D 157 TDNLVY-----RRLPNKGDP-SAE-----LD 176

QY 236 VPADSLITERPPNKEIVRYFIRHNALRAGTCENAPWVEDELVKYSLPSKFSDFLLDP 295
D 177 LPFROLRRSRMEFNLNLSMFTKSNVSRV-----DGLLR--PKPEAYKQVYVTP 223

QY 296 YKYMTLNPSTKRKNTGSPDRKPSKSKTDNSLSPLNPKLVCHVHLKSLSGSPKLVKN 355
D 224 ----GVNFST-----IFIGMPRYSPAKIK- 244

QY 356 SKNKSPEEHLEEMKOMSPNKLHNTFHI PKGPPAKPGKISDKPLKAKGSKGILNQ 415
D 245 -----KPDGK-----KQSTLNKYIVAGE 262

QY 416 KSTGSKSPKGLKTPKTKMQMTLLDMAKGTKQKTRAPRNSGGTPTSSKPKHLPPAA 475
D 263 ATAASKAKAK-----SDAKSLAELELV----- 286

QY 476 LHLIAYYKENDREDKRSALSCVISTARLLSSEDRARLPEELSLVOKRYELLEHKRW 535
D 274 -----SDAKSLAELELV----- 286

QY 536 ASMSEQRKEYLKKRBEIKKKLKKERREKEMLERLEKQ-----KRYBQELTG-KN 589
D 287 -----KREKEAKLIELEKQAEKKAQLIERVENECLNLLQTDLERTDQKV 333

QY 590 LPAELVDTPEGLNPTLFGDVAMVVEFLSCYSGLL--LPDAQYPITAVSLMEALSA-DKG 646
D 334 LPRYQIVTL--LPEHLIGDAFMKREFMHTYTGLLSGIEVFQRNLSFYEMTRALTAREJA 391

QY 647 GFLVLRVILLQTL-----QDEIAEDY-----GELG 675
D 392 G--PLSDILLVLLGTVPDLQKEEBEACAVYLDRAAQTOEPYWSMAQAASHLYAKRHFS 449

QY 676 MKLSEIPLTLHVSSELVRLCLRRSD--VOESEG-----SDTDDNKDSAAFEDNEVDQDF 728
D 450 FKVNELPDALTSLSEVLRLHLIGSGAFVNEKAERWRMYRNGYSKSGDPPGLELRLEHSHI 509

QY 729 LEKLETSEFFELTSEEKLIQILTALCHRLIMYTS-VQDHMETRQOQSAELERLAVLKEE 787
D 510 LRLKNHVSYQLKPKDINLLIRCLMSQI-MTYSGTINLIEERMEQTAKARQDLRALVGE 568

QY 788 NDKRAEKQKEKEMAKNKENGKVENGLGKTDKRRKRVKPEPQVDTEAEDM-ISA VKSR 846
D 569 NKRLAAVEINRKL-----TQMHLVNGVEPEK 597

QY 847 LLATQAKKERIQEREMKVKLERQAEERIRGHKAAAKAFQEGIAKAKLVNRRPTIGVD 906
D 598 REALVEKLKKGIAE-----LHAQSDQOH-RKHELQMLK-----LHSQLFNFLVYLGMD 644

QY 907 RNHNRYWLFSDVEVPGLFIEKGWVHDSIDYRPNHHCCKDHTVS----- 947
D 645 RCYRKYVYL-ESMPGIFVEHS--PDSLD-----TCLEQPIITNKSQIETRQOSALPNRKD 696

PR 01-MAR-2002; 2002US-00087192.
 XX (SAGR-) SAGRES DISCOVERY.
 XX Morris DW;
 XX WPI; 2003-328604/31.
 XX Recombinant nucleic acid useful for diagnosis and treatment of carcinoma
 PT comprises a nucleotide sequence.
 XX Claim 5; SEQ ID NO 663; Opp; English.
 XX The present invention relates to novel DNA and protein sequences which
 CC are associated with carcinomas. The sequences are useful for: (i) for
 CC screening drug candidates; (ii) for screening of bioactive agent capable
 CC of binding to Carcinoma Associated Protein (CAP); (iii) for screening of
 CC a bioactive agent capable of modulating the activity of CAP; (iv) for
 CC evaluating the effect of a candidate carcinoma drug; (v) for diagnosing
 CC carcinoma; (vi) for inhibiting the activity of CAP; (vii) for treating
 CC carcinoma; (viii) for neutralizing the effect of CAP; (ix) as a biochip;
 CC (x) for diagnosing carcinoma or a propensity to carcinoma; and (xi) for
 CC determining Carcinoma Associated (CA) gene copy number. In addition, the
 CC CA genes are useful as DNA vaccines and the CAP are useful as markers of
 CC carcinoma including lymphoma. The present sequence is one such CAP. Note:
 CC This patent is an equivalent to basic patent US2002182586A1, for which no
 CC sequence data was published
 XX Sequence 1586 AA;
 SQ
 Query Match 7.6%; Score 605; DB 7; Length 1586;
 Best Local Similarity 20.8%; Pred. No. 3.6e-31;
 Matches 315; Conservative 232; Mismatches 556; Indels 408; Gaps 59;
 QY 137 IHPLEKVDSEATE--KSDGACDPSDDKE-----NSSQIADHQKKEVYKE 182
 DB 236 IHPDEAAEKELSVVAENGTGLVLSLEEEQPELKMCGYNGSVSVESLHGOEVSVLVDP 295
 QY 183 -----DEGRRESINDRARRSPKLTSLKGERKWAPPKFLPHKY---DVKLQN-EDKI 232
 DB 296 PTVSCLDD--PSHLPDQEDTFLSSEDSLEPDSLAAPVSGSLYGIDDAELMGAEKDL 353
 QY 233 -----ISNVPAQLIRTERPPNKEIVRYFIRHNALRAGTGENAPWVEDE 277
 DB 354 PLEGNPVISALDPCALSNANAFSLADDQTSASIF-----VSPTSPPVLGES 401
 QY 278 LVKYSLSLPSDFLDLPKYMTLPSTKRYKTGSDPRKPSKSKTDNSLSPLNPKLW 337
 DB 402 VLQGSPEPSPAAAF-----QTVSPA--RKVVSAAPKARADRETTGGAVA----- 444
 QY 338 CHVHLKKSLSGSPKLVKNSKNSKSPHELEEMMKMSPNKLHTNPHIPK----- 386
 DB 445 -----VSGSGDVLK---RRATPEE-----VRPLQHGWRREVRIKGSRWQGETWY 489
 QY 387 KGPPAKPKGKSDKPLKAKGRSGKILNGQKSTGNSKSPKKGL-----KTPK-TMKQM 438
 DB 490 YGFCGKRMKQF--DEVIKYSRNVVHSVRREHFSFSPMPVGDFFFEEDTPEGLQWQL 546
 QY 439 TLLDMAKTQKTRAPRNSGGTPTRTSSK-PHGHLP-----PAALHIIAYKENKORE 489
 DB 547 SAEIEPSRTOAIT-----GKRGRPRNNEKAKNKEVPKVRGRGRPPPKIKMPELNTDNRL 602
 QY 490 DKRSALSVCISKTARLLSSEDRARLPEELRSVLQKRYELLEHKRWASMSSEQRKEYLKK 549
 DB 603 PKK-----LETQILSDEDDKAKTKNKKVQRQK-VQRGESQTPVQOQARNKRQDTPKS 654
 QY 550 -KREELKKLKBKAKE-RREKEMLERLEKRYEQDELTG-KNLPAPRLVDTPEG--LPN 604
 DB 655 LKQKDTKKLKKRLEEQQQQAILE--EMKKPTEDMCLSDHQLPDPFTRI---PGLTSS 709
 QY 605 TLFQGVAMVVEFLSCYSGLLLPDAQYPITAVS-LMEALSADKGGFLYLNRLVILLQTL 663
 DB 710 RAFPSCLTIVFELHSGKVLGFLDKVDPSLGLVQEGLLCOGDSLDKVDQLLVRLKAL 769

RESULT 36

ABM85417

ID ABM85417 standard; protein; 1873 AA.

XX ABM85417;

AC ABM85417;

XX DT 18-NOV-2004 (first entry)

QY 664 QDEIAEDYGE-----LGMKLSIPLTHSVSELVRLCLRRSDVQEESEGSDDTDNDKDSAAF 719
 DB 770 HDPGLPPYCQSLKILGKMXSEIPLTRDNVSEILRCFLMA----- 808
 QY 720 EDNEVQDEFLEKLETSEFPFELTSEKLIQILTALCHRI-----LMTYSVQDHMETRQMSAE 775
 DB 809 --YRVEPSFCDLSRTQPPQAPQPOQKAAILAFVLHFLNSSTIIINBIDKLTLESVSSCKRN 866
 QY 776 LW--KERLAVLKEENDKKRAEKQKKEKMEAKNGKVENGLGKTDRKRIKVFEBQV-D 832
 DB 867 KWIIVEGRLLRLTALAKRTGRPEVME-----GAEDGLQRR-RSRRIEETSIGIEE 916
 QY 833 TEAEDMI SAVKSRRLAIQAQKKEIOEREMKV-KLEROAEEERIRKHAAAAPQEBGI 891
 DB 917 EEEENTTAVHGRR-----GRKEGIDVAASSIPELERHIEKLSKR-----QLFPRKKL 965
 QY 892 AKAKLVMRTPTGTDNRNHRNYLWLFSDVEPGLFIE--KGWVHDSIDYRFNHC-----KDH 944
 DB 966 LHSSQMLRAVSLGQDRYRRHYWL-PYLAGIFVEGSEGSTGPMFNFSAQIPCDAAPTPPP 1024
 QY 945 TVSGDEDVCPRSKKNALGNASMTN----- 969
 DB 1025 AVSEDQ-----PTPSLQLLASSKPMNTPGAANPCSPVOLSTHPLPGGTPKRLSGDSEMSQS 1081
 QY 970 -----QHGTATEVAVETTPKQGNLWFLCDSQKELDELLNCL 1007
 DB 1082 PTGLQPKRGRPPSKFFKQVEQH-YLTQLTAQPIPPCMSCGMMWRDPE-TLDVLLKAL 1139
 QY 1008 HPQIRESQLKERLEKRYODIHSIHLARKPNGLKSCDGNQELLN--FLRSDLIEVATR 1065
 DB 1140 HPRGIREKALHKLSK-HKDFLOEYCL-----QPLTDPIFEPNEL-----PA 1180
 QY 1066 LQKGLGVVEETSEFEARVISLEKLDGCECVIALQASVKKFLOGFMAP----- 1115
 DB 1181 LEEGVMSPKKTYTETDLAVLQWVEELEQRVLSDLQI-----RGWTCPTDPTDSTREDLT 1235
 QY 1116 -----KQKRRLQSEDSAKTEE 1132
 DB 1236 YCEHLPDSPEDIPMRGRREGTVPQQRNNPLDLAVMLLAVLEQNVERRYLREPLWAAHE 1295
 QY 1133 VDEEKQKVE-----EAKVASALEKWKTAIRAQAQFTRSMHVLLGLMDACIKWD 1179
 DB 1296 VVVERKALLSTPNGAPDGTSTEISYBITPRVVRWROTLCRCSAAQVCLCMGOLERSIAWE 1355
 QY 1180 MSAENARCKVCPKKGDDKLIILCDSCNAFLHFLCLRPALYEPVDGEWQCPACQATARN 1239
 DB 1356 KSVNKTCLVCKRGNDLFLCCDGRGCHYCHRPKMEAVPEGWDFCAVCL----- 1408
 QY 1240 SRGRNYTESASEDSEDEDEEBEEDYEVAGLRRLRPRKTTIRGKHSVIP--- 1296
 DB 1409 -----SQVEEYTORFGFPKRGOK-----RKSSFPLTF 1436
 QY 1297 PAARGRPCKKPHSTRSQKAP--PVDDAEVDELVLQTKSSRRSQSLQKCEBILHK 1354
 DB 1437 PEGDSRRR-----MLSRSDSPAVPRYPEDGLSPPK---RRRHSRSHHSDLTFCETILME 1489
 QY 1355 IVKYRFSPPRPVTRDEADYVDVITHPMDFTQVNCSCGYSYVSQSFELTDMKQVFTN 1414
 DB 1490 MESHDAAMPFLFEPVNPRLVSGYRRVKNPMDFTMRERLLRGYTSSEBFAADALLVFDN 1549
 QY 1415 AEVYNCRGSHV 1425
 DB 1550 CQTFNEDDSEV 1560

XX Human protein sequence hCP42010.
DE Cytostatic; carcinoma; lymphoma; cancer; human.
KW Homo sapiens.
OS
XX WO2003073826-A2.
FN 12-SEP-2003.
PD
XX 28-FEB-2003; 2003WO-US006235.
PF
XX 01-MAR-2002; 2002US-00087192.
PR
XX (SAGR-) SAGRES DISCOVERY.
PA Morris DW;
XX
XX WPI; 2003-328604/31.
DR
XX Recombinant nucleic acid useful for diagnosis and treatment of carcinoma
PT comprises a nucleotide sequence.
XX
XX Claim 5; SEQ ID NO 666; Opp; English.
XX
XX The present invention relates to novel DNA and protein sequences which
CC are associated with carcinomas. The sequences are useful for: (i) for
CC screening drug candidates; (ii) for screening of bioactive agent capable
CC of binding to Carcinoma Associated Protein (CAP); (iii) for screening of
CC a bioactive agent capable of modulating the activity of CAP; (iv) for
CC evaluating the effect of a candidate carcinoma drug; (v) for diagnosing
CC carcinoma; (vi) for inhibiting the activity of CAP; (vii) for treating
CC carcinoma; (viii) for neutralizing the effect of CAP; (ix) as a biochip;
CC (x) for diagnosing carcinoma or a propensity to carcinoma; and (xi) for
CC determining Carcinoma Associated (CA) gene copy number. In addition, the
CC CA genes are useful as DNA vaccines and the CAP are useful as markers of
CC carcinoma including lymphoma. The present sequence is one such CAP. Note:
CC This patent is an equivalent to basic patent US2002182586A1, for which no
CC sequence data was published
XX
SQ Sequence 1873 AA;

Query Match 6.8%; Score 540.5; DB 7; Length 1873;
Best Local Similarity 19.3%; Pred. No. 9.9e-27;
Matches 263; Conservative 183; Mismatches 432; Indels 487; Gaps 41;

QY 415 QKSTGNSPKKGLTPKTKMQLTLLDWAKGTQ---KMTAPRNSGGTPTRTSSKPHKHL 471
DB 616 QAITGKGRPR---NTEAKTYEVPKVKRGRPRPKVITEL-----LNKTDNRPLKLL 666
QY 472 PPAALHLIAYYKENKDRDKRSALSVCVSKTAR-----LLSSEDRARLPEEL 518
DB 667 EA-----QETLNEEDKAK-----IAKSKKMRQKVQGECCQTITQGGARNKKQET 712
QY 519 RSLVOKRYELLEHKRWASMSSEQRKYLKKREELKKLKE---KAK----- 563
DB 713 KSLQKQKAKKSKKKGKTKQKELKVKRKEKKEKVKMKKEEVTAKKAPACKADKTLATQ 772
QY 564 ---ERREKEMLERLEKQRYEDQELTG-KNLPAPRLVDTPTEG--LPNTLFGDVAMVVEFL 617
DB 773 RLREERQKQMLEEMKQPTEDMCLTDHPDFSRV---PGLTLPFGAFSDCLTIVEFL 829
QY 618 SCYGLGLPDAAQYPIYAVS-LMEALSADKGGFLYLNRLVILLOTLLQDEIAEDYGE--- 673
DB 830 HSFQKVLGFDPAKQVPSLVQLQEGQLCQDLSLGEVQDILLVLLKALHDPGPPSYCQSLK 889
QY 674 -LGMKLSIPLTLHSVSELVRLCLRRSDVQBESESGDTDNKKDSAAFEDNEVQDFLEKL 732
DB 890 ILGEKVSIPILTRDNVSBILRCFLMAYG-----VEPALCDRL 926
QY 733 ETSEPFELTSEBKLIITLALCHRI-----LMTYSVDHMETRQMSAELW--KERLAVLKE 786

DB 927 RTQFPQAPPOQKAAVLAFVLHVLNGSTLIINEIDKTLTLESMSYKKNKWIVVEGLRRLK- 985
QY 787 ENDKKRAEKQKRWEMAKNKGKVENGLGKTDKRRIVKFEQVDTAEADMISAVKRR 846
DB 986 ---TVLAKRTGRSEVEMEGP-----EECLGRRRSRIMEETSGMBEEEEESIAAPVGR 1037
QY 847 LLAIQAKKERIQEREMKV-KLREQAEERIRKHKAAAEKAFQEGIAKAKLVMMRTPIGT 905
DB 1038 -----GRDGEVDATASSIPELEQIEKLSK-----QLFRKKLLHSSQMLRAVSLQ 1086
QY 906 DRNENRYWLFSDVPGLFIE-----KGMVHDSIDYRF----- 937
DB 1087 DRYRRRYWVL-PYLAGIFVEGTGNLVPDEVIKKTETSLKVAHAASLNPALFPMKMLAG 1145
QY 938 -----NHHCKDHTVSGDEDYCPRSK----- 957
DB 1146 SNTTASSPARARGPRKTKPGSMQPRHLKS-PVRGQDSEQPQALQPEAQLHAPAQPOFQ 1204
QY 958 -----KANLGKN-----ASMNTQHGTTATEVAVETTTTPKQG- 987
DB 1205 LQLQLQSHKHGFLQEGSPLSGQSHDLSQAFSLWSLQTSQSHSLSSVLTPDSSPKG 1264
QY 988 -----QNLWFLCDSQ----- 997
DB 1265 LDPAPSQPPEPEDEAESSDPQALWFNISAQMPCNAAPTPPPAVSEDQTPSPQQLAS 1324
QY 998 ----- 997
DB 1325 SKPMNPSAANPCSPVQFSSTPLAGLAPKRRAGDPGEMPOSPTGLGQPKRRGRPPSKFFK 1384
QY 998 -----KELDELNCLHPQGIRESOLKERLEKRYODI 1028
DB 1385 QMEQRYLTQLTAQVPPPEMCGWWIRDPEDLWDLKALHPRGIREKALHKLHNK-HRDP 1443
QY 1029 IHSIHL--ARKPNLGLKSCDGNQE-LNLFRLSDLIETATRLKQGLGYVEETSEFEARVI 1085
DB 1444 LQEVCLRSADPIFEPRQLPAFQSGIMSPKE-----KTYETDLAVLQWVEELEQORVI 1497
QY 1086 S-----LEKLKDFQECV----- 1097
DB 1498 MSDLIQIRGWTCPSPDSTREDLAYCEHLSDSQEDITWRGRGREGAPQRKTTNPLDLAVMR 1557
QY 1098 -IALQASVIKKFLOGFMAPKQRRKIQSEDSAKTEEVEDEKKQVBE-----EA 1143
DB 1558 LAALQONVERKYLREPLW-----THEVVLEKALISTPNGAPEGTTEISY 1603
QY 1144 KVASALEKWKTAIREAQTFSRMHVLLGMLDACIKWMSAENARCKVCPPKGGEDDKLILCD 1203
DB 1604 EITPRIVRWQTLERCPSAAQVCLCGLQLERSIAWEKSVNKVTCILVCRKGDNDDEFLLED 1663
QY 1204 ECNKAFHLFCLRPALYEPVDPGEWQCPACQATARNRGRNYTBESASEDESDDEE 1263
DB 1664 GCDRGCHYCHRPKMEAVPEGDWFCV-----LAQVQVEGEFTQPG 1705
QY 1264 EEEEEEEEDYEV---AGLRLRPRKTIKRGHSHVIPPAAARSGRRRPGKPHSTRSQPKAP 1320
DB 1706 FPKGQKKSQSYSLNFSEGGRRRRVLRGRES--PAAGPRYSBEGLSPSKRR----- 1757
QY 1321 PVDDAEVDLVLQKRSRRQSLQKCEELHKHVIKRYFSWFPFPREPTVTRDEAEDYDVI 1380
DB 1758 -----LSMRNHHSDLTFCCEILMEMESHDAWPFLEPNVPLVSGYRRII 1802
QY 1381 THPMDFTQVQNKCGSYRSYRQVEFTDMKQVFTNAEVNCRGSHV 1425
DB 1803 KNPMDFSTMRRLRLLGGYTSSEFAADALLVFDNCQTFNEDDSEV 1847

RESULT 37
AAB95554
ID AAB95554 standard; protein; 1727 AA.
XX
AC AAB95554;
XX

QY 1309 PHSTERSQKAPPVDDAEVDELVLTKSSRRSQSLELOKCEILHKIVKYRFSWPPRPV 1368
Db 1606 PSKRR-----LSMRNHSDDLTCFCIILMEMESHDAAMPFLSPV 1644
QY 1369 TRDEAEDYDVITHPMDFQTVQNKSCGYSRQVQFLTDKQVFNAAVYNCRGSHV 1425
Db 1645 NPRLVSGYRRIKKNPWFSTWRERLLRGYTSSEFAADALLVFNCTCFNEDDSEV 1701
RESULT 38
ID AAM40239
ID AAM40239 standard; protein; 1878 AA.
XX AAM40239;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polypeptide SEQ ID NO 3384.
XX
KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokineic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia.
XX
OS Homo sapiens.
XX
PN WO200153312-A1.
XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000WO-US034263.
XX
PR 23-DEC-1999; 99US-00471275.
PR 21-JAN-2000; 2000US-00488725.
PR 25-APR-2000; 2000US-00552317.
PR 20-JUN-2000; 2000US-00598042.
PR 19-JUL-2000; 2000US-00620312.
PR 03-AUG-2000; 2000US-00653450.
PR 14-SEP-2000; 2000US-00662191.
PR 19-OCT-2000; 2000US-00693036.
PR 29-NOV-2000; 2000US-00727344.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Auendi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;
PI Zhou P, Goodrich R, Drmanac RT;
XX
DR WPI; 2001-442253/47.
DR N-PSDB; AAI59395.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders such
PT as central nervous system injuries.
XX
PS Example 5; SEQ ID NO 3384; 10078pp; English.
XX
CC The invention relates to human nucleic acids (AAI57798-AAI61369) and the
CC encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activ/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders. Note: The sequence data for this patent did not form

CC part of the printed specification

XX Sequence 1878 AA;
SQ
Query Match
Best Local Similarity 6.7%; Score 537.5; DB 4; Length 1878;
Matches 269; Conservative 178; Mismatches 421; Indels 509; Gaps 44;
QY 415 QKSTGNSKPKGLKTKTKMKTLLDMAKGTQ---KMTAPRNSGGTPTRTSSPHKHL 471
Db 619 QAITGKGRPR---NTEKATKEVPKVKRGRPPKVKITEL-----LNKTDNRPLKL 669
QY 472 PPAALHLIAYKENKORDEKESALSVCVSKTAR-----LSSSEDRARLPREL 518
Db 670 EA-----OETLNEDKAK-----IAKSKKMKRQKVQGECCOTTIQQARNKRKQET 715
QY 519 RSLVQRYELLEHKRWMASMEEQREKYLEKKRBELEKKLKEKAKER---REKEML---- 571
Db 716 KSLAQK-----EAKKK-----SKAEKEGKTQKEKLEKVKREKEKVKMKEKEVTKAK 765
QY 572 -----ERLEKOKRY-----EQELTG-KNLPAPRLVDTPEG--LPNT 605
Db 766 PACKADKTLATQRRLLEERQRMILEEMKKTEDMCLTDHQPDPFSRV---PGLTLPFG 822
QY 606 LFGDVAMVVEFLSCYSGLLLPDAQYPTAVS-LMEALSADKGGFLYANRVLVILLQTLQ 664
Db 823 AFSCLTIVFHLHSGVIGFDPKDVPSLVGLQEGLLCOGDSLGEVODLLVRLKAAHL 882
QY 665 DEIAEDYGE---LGMKLSIPLTIHVSSELVRLCLRRSDVQSESESGDSTDNDKDSAAFE 720
Db 883 DPGFPSCQSLKILGEKVSEIPLTRDNVSEILRCFLMAYG----- 922
QY 721 DNEVDFLEKLETSEPFELTSEKLIQILTALCHRI-----LMTYSVQDHMETRQMSAEL 776
Db 923 ---VEPALCDRLRTQPPQAPPPQAAVLAFLVHENGSTLIINEIDKTESMSSYRNK 979
QY 777 W--XERLAVLKEENDKKRAEKQKBEAKNKENGKGVENGKGTDRKKRIKVFEPQVDE 834
Db 980 WIVGRLRRLK---TVLAKRTGRSEVMEGP-----EECLGRRRSSRIMETSQWEEEE 1030
QY 835 AEDMISAVKSRLLAIQAKEREIQEREMKV-KLERQAEERIRKHAAKAFAQEGIAK 893
Db 1031 EESIAAVPGR---GRRDGEVDATASSIPELERQIEKLSK-----QLFFRKLLH 1079
QY 894 AKLVNRTPIGTDRNHNRYWLFSDVEVPGLPTE-----KGWVHDSID 934
Db 1080 SSQMLRAVSLGQDRYRRYYVL-PYLAGIFVEGTEGNLVPEVVIKKTDSLKVAHAASLN 1138
QY 935 YRF-----NHHCKDHTVSGDEDCYCRSK-----ASMNTQHTGATE 976
Db 1139 PALFSMKMELAGSNNTASSPARARPRKTKPGSMQPHLKS-PVRGQDSEQPQALOPE 1197
QY 958 -----KANLGN----- 976
Db 1198 AQLHAPAPQPOLQLOLQSHKGFLEQSGPLSGSQHDLQSQAFSLMSLQTSQSHSLLS 1257
QY 977 VAVETTTPKQG-----QNLWFLCDSQ----- 997
Db 1258 SSVLTPDSSPGKLOPAPSPQPEPEPEDEAESSPPQALWFNISAQMPCAAPTPPAVSE 1317
QY 998 ----- 997
Db 1318 DOPTSPQOLASSKPMNRPSPANPCSPVQFSTPLAGLAPKRRAGDGPMPQSPGLGQP 1377
QY 998 -----KELDELNCLHPQGIRESQ 1016
Db 1378 KRGRRPSPKFFKQEQRYLTQITAPVPPVPMCSGWWIRDPEDMLDAMLKALHPRGIREKA 1437
QY 1017 LKERLEKRYQDIHSIHU--ARKNWLGLKSCDGNQE-LLNFLRSLDIEVATRLQGGIGY 1073
Db 1438 LHKHLNK-HRDFLOEVCVLRPSADPIFEPRQLPAFQEGIMSWSPKE-----KTYETDLAV 1490
QY 1074 VEETSEFEARVIS-----LEKLDKDFGECV----- 1097

Db 1491 LQWVELEQVIMSDIQRGWTCFSPDSTREDLAYCEHLSDSQEDITWRGRGREGIAPQR 1550
Qy 1098 -----IALQASVIKFLQGFMAPKQKRKLQSEDSAKTEEVEDEEKWVE--- 1141
Db 1551 KTNPLDLAVMLLALEQNVRYLREPLW-----THEVLEKALLSTPN 1596
Qy 1142 -----EAKVASALEKWKTAIREAQTFSRMHVLLGLMDACIKWDMVSAENARCKVCP 1191
Db 1597 GAPEGTTTSEIYETPRIRVWRQLERCRSAQVCLGLQLERSIAWKSVMKVTCLVCR 1656
Qy 1192 KGEDDKLILCDCKNAFLCLRPALYVDFGECWCPACOPATARNRGRNYTESAS 1251
Db 1657 KGNDFLLCLDCDGRGCHYCRPVEAVPEGDWFTVC-----LA 1698
Qy 1252 ESEDEDESEEEEEEEEDYEV---AGLRLRPKTRIRGHSHVPPAARSGRRPGKK 1308
Db 1699 QQVEGEFTQKPGPKQKQKSGYSLNFSEGDGRRRRVLLRGES--PAAGPRYSEGLS 1756
Qy 1309 PHSTRSQPKAPPVDDAEVDELVLQTKRSSRROSLELQKCEELHKKIVKYRFSWPPRPV 1368
Db 1757 PSKRRR-----LSMRNHSDLTFCEILMEMESHDAWPPLEPV 1795
Qy 1369 TRDEADYDVITHPMDFQTVQNKSCGYSYRVSQVQFELTDMKQVFTNAEVNCRGSHV 1425
Db 1796 NPRLVSGYRIIKNPMDFSTMRERLLRGGYTSSEEPFAADALLVFDNCQTFNEDDSEV 1852

RESULT 39
AAM42025
ID AAM42025 standard; protein; 1589 AA.
AC AAM42025;
DT 22-OCT-2001 (first entry)
DE Human polypeptide SEQ ID NO 6956.
KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia.
KW Homo sapiens.
OS
PN WO200153312-A1.
PD 26-JUL-2001.
PF 26-DEC-2000; 2000WO-US034263.
PR 23-DEC-1999; 99US-00471275.
PR 21-JAN-2000; 2000US-00488725.
PR 25-APR-2000; 2000US-00552317.
PR 20-JUN-2000; 2000US-00598042.
PR 19-JUL-2000; 2000US-00620312.
PR 03-AUG-2000; 2000US-00653450.
PR 14-SEP-2000; 2000US-00662191.
PR 19-OCT-2000; 2000US-00693036.
PR 29-NOV-2000; 2000US-00727344.
XX (HYSE-) HYSEQ INC.
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Zhou J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;
PI Zhou P, Goodrich R, Drmanac RT;
XX WPI: 2001-442253/47.
DR N-PSDB; AA161181.
XX Novel nucleic acids and polypeptides, useful for treating disorders such

as central nervous system injuries.
Example 2; SEQ ID NO 6956; 10078pp; English.
CC The invention relates to human nucleic acids (AA157798-AA161369) and the
CC encoded polypeptides (AA38642-AA42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemia and
CC C.N.S disorders. Note: The sequence data for this patent did not form
CC part of the printed specification
XX
SQ Sequence 1589 AA;
Query Match 6.7%; Score 536; DB 4; Length 1589;
Best Local Similarity 19.4%; Pred. No. 1.6e-26;
Matches 281; Conservative 185; Mismatches 430; Indels 554; Gaps 48;
Qy 415 QKSTGSKSPKGLKTPKTKMQMTLLDMAKGTQ---KMTAPRNSGGTPTSSKPKHKL 471
Db 288 QAITKGRGRPR---NTEKARTKEVPKVGRRGPPKVKITEL-----LNKTDNRRLPKKL 338
Qy 472 PPAALHLIAYYKENDREDKRSALSQVSKTAR-----LLSEDRARLPBEL 518
Db 339 EA-----QETLNEEDKAK-----IAKSKMKRQKVQRGBECQTTIQGAARNKKEQT 384
Qy 519 RSLVKRVYELHKKRWASMSSEQRKEYLKKRBEELKKLKKAKAKER---RKEML----- 571
Db 385 KSLKQK-----BAKKK-----SKAEKKGKTKQEKKEKVKREKKEKVKRKEEYTKAK 434
Qy 572 -----ERLEKOKRY-----EDQELTG-KNLPAFLVDTPEG--LPNT 605
Db 435 PACKADKTLATORRLEERQRMILEEMKKPTEDMCLTDHQPDPFSRV---PGLTLPSPG 491
Qy 606 LFGDVAMVVEFLSCYSGLLLPDAQYPTAVS-LMEALSADKGGFLYLNRLVILLQTLQ 664
Db 492 AFSDCLTIVEFLHSFGVLGDPDPAKDVPSGLVGLQGLQCGSDLSGEVQDLLVRLKAAHL 551
Qy 665 DEIAEDYGE---LGMKLSSEPLTLHVSVELVRLCLRRSDVOEESGSDTDNKGSAPE 720
Db 552 DPGFSPYCSQSLKILGEKVSIEPLTRDNVSEILRCLMAYG----- 591
Qy 721 DNEVQDEFLEKLETSEFFELTSEEKLQILTALCHRI---LMTYSVDHMETRQMSAEL 776
Db 592 ---VEPALCDRLRTOPFOAQPPQKAAAVLAFVLHENGSTLIINEIDKLTSSMSYRKNK 648
Qy 777 W--KERLAVLKEENDKDAEKOKRMEAKNKENGKVENGLGKTDRKKRIVKFEPOV-DT 833
Db 649 WIVEGRLRLK---TVLAKRTGRSEVMEGP-----SECLGR-RSSIMEVTSQMESE 698
Qy 834 EAEDMISAVKSRRLIAIOAKKERETOEREMKV-KLERQAEERIRKHAASKAFOEGTA 892
Db 699 EEEESTAAVPGRR-----GRRDGEVDATASSIPELERQIEKLSKR-----QLFPRKLL 747
Qy 893 KAKLVMRTPPTGTDNRNRYMLFSDVPCLFTE-----KGWVHDSI 933
Db 748 HSSQMLRAVSLGQDRYRRYVWL-PYLAGIFVEGTEGNLVPEVTKETDSLUKVAHASL 806
Qy 934 DYRF-----NHHCKDHTVSGDEDYCPRSK--- 957
Db 807 NPALFSMKMELAGSNTTASSPARAGRPRKTKPGSMQPRHLKS-PVRGDSQPQAQLQP 865
Qy 958 -----KANLGN-----ASNTOHGTTAT 975
Db 866 EAQLHAPQPOPQLQLQSHKGFLEQSGSPLSLGQSQHDLSQSAPLSWLSQTSQSHSLL 925

Db 661 REBSRMRKGRPPNVNDAEFLDNADAKLLRLKLOAQEIARQAAQIKLLRKLQKQEOARVA 720
Qy 422 KSPFK--GLKTPKTKOMQTLDMAGTKQMTAPRNSGGTPTRTSSKPHKHLPPAALHLI 479
Db 721 KEAKQQAIAWAEERKQKEQIKIMQKQEKIRIQO-----IRMEKEL--RAQQL 769
Qy 480 AYYKENKDREDKRSALSCVISKTARLLSSBEDARLPEELRSLVKRYELLEHHKRWASMS 539
Db 770 EAKKKKKE-----ANAKULAEKRIK-EKEMR--RQAVLLKHQER----- 809
Qy 540 EQRKEYLKKREELKKLKAERREKEMLEKQKRYEDOEL--TGKNL----- 590
Db 810 ERRQHMLMKAEARKKAEERLQKQKDEKRLNKERLKEORRLEMAKELKKNED 869
Qy 591 -----PAFLVDTPPE-GLPNTLFGDVAMVVEFLSCYSGLLPDAQYPIITAVSLMEALS 642
Db 870 MCLADQKPLPELPRIGLVLSGSTFSDCLMVQVFLNFGKVLGFDVNDVNLVQLQGL 929
Qy 643 ADKGGFL-VLNRVLVILLQTLQDE-IAEDY---GELGMKLSEIPIUTLHSVSELVRLCLUR 697
Db 930 LNIQDSMGVEVDLLVRLLSAAVCDPLITGYKAKTALGEHLLNVGNRDNVSEILQIFW- 988
Qy 698 RSDVQSESGSDTDNDKSAAPEDNEVQDEFLEKLETSEFFELTSEKQLILTALCHRL 757
Db 989 -----EAHCGQTELTESLTKAFQAHTPAQKASVLAFLINELA 1026
Qy 758 MTSYVDHMETRQMSAELWKERLAVLKEENDKRAEKQKREMEAKNKGKVGNG--- 814
Db 1027 CSKSVSEIDKNIDYNSLRDKWV---EGKLRLRIIHAKTGKRTSGIDLGEEQ 1082
Qy 815 --LG--KTRDKRIRVKFEPQVDTEADNMISAVKSRLLAIQAKKERRIQE--REMK 864
Db 1083 HPLGTPTPGRKRRKGGSDYDDDDDDSD-----QDEDEDEBEDKQK 1129
Qy 865 VKLERQAEERIRKHKAAEK-----AFQEGIAKAKLVMRTPITGTDNRNR 911
Db 1130 GKRTDICEDEBGDQAASVEELEKQIEKLSKQOQYRRKLFDASHLSLRVSMFGDPRYRR 1189
Qy 912 YWLFSDVPGLFTE-----KGWVH-----DSIDYRFNHC 941
Db 1190 YML-PRCGGIFVEGMESEGLEEIAKEREKLLKAESVQIKEMPETSGDSLNCSTHIC 1248
Qy 942 -----KDHT---VSGDEDYCPRSK-----KANLGKNA----- 965
Db 1249 EQKEDLKEKXNDNLFQKPGFSKLSKLEVAKMPPESEVMTPKPNAGANGCTLSYQNSG 1308
Qy 966 --SMNTOHGTATEVAE-----TTTTPKOQNLWFL- 993
Db 1309 KHSLSGVQSTATQSNVEKADSNLFWTSGSGPKGFYSPLPNDQLLTKLTKNRQ--WFSL 1366
Qy 994 -----CD----- 995
Db 1367 LPRTPCDDTSLTHADMSTASLVTPOSPKSPSPTPAPLGSSAQNPVGLNPFALSPLOV 1426
Qy 996 ----- 995
Db 1427 KGGVSMGLQFCGWPTGVVTSNIPFTLSVPSLGSGLGSEGNNGSFLTSNVASKSESPV 1486
Qy 996 -----SOKELDELINCLHPQIRES 1015
Db 1487 PONEKATSAQAANAEEVAKVPDPFPSPKPIPEMQFGWRIIDEDLKALKLVHLRGIREK 1546
Qy 1016 QKLERLEKRYQDI IHSIHLARKPNLGLKSCDGNQELLNFRSLDIE---VATRLQKGLG 1072
Db 1547 ALQKIQK-HLYIYIQAQCLNKK-DVNAIELNEER--NOVTRDIVENWSVEEQAMEMDLS 1602
Qy 1073 YVEETSEFARVISLEKLDKDFGECVIALQASVIKFKLQGFMAPKQ----- 1117
Db 1603 VLQOVEDLERRVAS-----ASLQ-----VKGMCMPEPASEREDLVYFEHKSF 1644
Qy 1118 -----KRRKLQSEDSAKTEVDEEKKMVE--EAKVASALEKWK 1153
Db 1645 TKLCKEHGDGFTGEDESSAHALERKSDNPLDIAVTRLADLERNIERRIEBEDIAPLRVWR 1704

Qy 1154 TAIREAQTFSRMHVLLGMLDACIKWDMSAENARCKVCPKKGEDDKLILDECNKAFLHFC 1213
Db 1705 RALSEARSAQVALCICQQLQKSIAMEKSTMKVYVCQICRKGDNELLLCDDGCKGCHTYC 1764
Qy 1214 LRPALYVDPDGWQWQACQAPATA-----RRNSGRNNTESASE-----DSEDESD 1260
Db 1765 HRPKITTPDGMFCFACIAKASGOTLKITKLHVKGKKTNESKKGKKVTLTGDTDEDS- 1823
Qy 1261 EEEEEEEEBEEDYEVAGRLRPRKTIRGKHSVIPPAARSGRRPGKPKPHSTRRSOPKAP 1320
Db 1824 -----ASTSSSLKRGNKDLQKQKMEENTS 1847
Qy 1321 PVDDAEVDELVLQTKRSRRRQSLQKCEIILHKIVKRFSPFPFPVTRDEADYDVI 1380
Db 1848 --INLSKQESFTSVKKPKRDDSKDLALCSMILTEMETHEDAWPFLLPVNLKLVPGYKVI 1905
Qy 1381 THPMDFQTVONKSCSGSVRSVOEFLTDMKQVFTNAEVYN 1419
Db 1906 KKPMPDFSTIREKLSGQYFNLETFAFDVRLVDFDNCETFN 1944

RESULT 41

ABR64241
ID ABR64241 standard; protein; 1972 AA.
XX AC ABR64241;
XX AC ABR64241;
DT 15-OCT-2003 (first entry)
XX X
DE Angiogenesis protein BNO150.
XX X
KW Cytostatic; antirheumatic; antiarthritic; antidiabetic; ophthalmological;
KW antipsoriatic; antiarteriosclerotic; cardiac; vasotropic; angiogenesis;
KW gene therapy; vasculature; cancer; rheumatoid arthritis; psoriasis;
KW diabetic retinopathy; cardiovascular disease; atherosclerosis;
KW ischemic limb disease; coronary artery disease.
XX X
OS Homo sapiens.
XX X
PN WO2003027285-A1.
XX X
PD 03-APR-2003.
XX X
PF 19-SEP-2002; 2002WO-AU001282.
XX X
PR 27-SEP-2001; 2001AU-00007973.
PR 27-SEP-2001; 2001AU-00007974.
PR 11-OCT-2001; 2001AU-00008210.
PR 29-OCT-2001; 2001AU-00008532.
PR 13-NOV-2001; 2001AU-00008838.
PR 28-AUG-2002; 2002AU-00951032.
XX X
PA (BION-) BIONOMICS LTD.

Gamble JR, Hahn CN, Vadas MA;

WPI: 2003-354655/33.

N-PSDB; ACF34516.

New angiogenic genes and polypeptides, useful for diagnosing,
prognosticating or treating an angiogenesis-related disorder, e.g.
cancer, rheumatoid arthritis, diabetic retinopathy, psoriasis or
cardiovascular diseases.

Claim 15; SEQ ID NO 176; 90pp; English.

The invention relates to the isolation of novel genes (ACF34446-ACF34559)
encoding proteins (ABR64180-ABR64281) involved in the process of
angiogenesis. The nucleic acid molecules are useful in identifying and/or
obtaining full-length human genes involved in an angiogenic process. The
nucleic acid molecule, polypeptides or complexes encoded, cells or
genetically modified non-human animals derived from these are useful for

Db 1848 --INLSKQSFSTSVKPKKDDSDKDLALCSMLTETHEDAMPFLPVNLKLVFGYKKVI 1905
QY 1381 THPMDFTQVONKSCGSYRSVQEFLLDMKQVFTNAEVYN 1419
Db 1906 KKPMDFSTIREKLSGQYFNLETFFALDVLVDFDNCETFN 1944

RESULT 42
ADP54420
ID ADP54420 standard; protein; 1972 AA.
XX ADP54420;
XX
DT 18-NOV-2004 (first entry)
XX Human PRO protein sequence SEQ ID NO:396.
XX human; PRO; immune related disease; inflammatory immune response;
KW immune response stimulation; antiallergic; antianaemic; antiarthritic;
KW antiasthmatic; antidiabetic; antiinflammatory; antipsoriatic;
KW antirheumatic; antithyroid; CNS; dermatological; gastrointestinal;
KW haemostatic; hepatotropic; immunostimulant; immunosuppressive; muscular;
KW nephrotropic; neuroprotective; osteopathic; respiratory; vasotropic;
KW virucide; gene therapy.
XX
XX Homo sapiens.
XX WO2004039956-A2.
XX
XX 13-MAY-2004.
XX
XX 28-OCT-2003; 2003WO-US034381.
XX
XX 29-OCT-2002; 2002US-0422472P.
XX
XX (GETH) GENENTECH INC.
XX
XX Aggarwal S, Clark H, Gurney AL, Schoenfeld J, Williams PM;
PI Wood WI, Wu TD;
XX
XX WPI: 2004-376182/35.
XX N-PSDB; ADP54419.
XX
XX New PRO polynucleotides and polypeptides, useful in diagnosing
PT and treating an immune related disease, e.g. systemic lupus
PT erythematous, rheumatoid arthritis, diabetes mellitus or asthma and in
PT stimulating an immune response.

PS Claim 1; SEQ ID NO 396; 3009pp; English.
XX
XX The present invention describes an isolated PRO nucleic acid (I). Also
CC described: (1) a vector comprising (I); (2) a host cell comprising the
CC vector of (1); (3) a process for producing a PRO polypeptides; (4) an
CC isolated PRO polypeptide; (5) a chimeric molecule comprising the
CC polypeptide of (4) fused to a heterologous amino acid sequence; (6) an
CC antibody which specifically binds to a polypeptide of (4); (7) a
CC composition of matter comprising a polypeptide of (4), an agonist or
CC antagonist of the polypeptide or an antibody that binds to the
CC polypeptide in combination with a carrier; (8) an article of manufacture
CC comprising a container, a label on the container and a composition of
CC matter of (7); (9) a method of treating an immune related disease in a
CC mammal; (10) a method for determining the presence of a PRO polypeptide
CC in a sample suspected of having the polypeptide; (11) a method of
CC diagnosing an immune related disease or an inflammatory immune response
CC in a mammal; (12) a method of identifying a compound that inhibits or
CC mimics the activity of or expression of a gene encoding a PRO polypeptide
CC ; and (13) a method of stimulating the immune response in a mammal. The
CC PRO sequences have antiallergic, antianaemic, antiarthritic,
CC antiasthmatic, antidiabetic, antiinflammatory, antipsoriatic,
CC antirheumatic, antithyroid, CNS, dermatological, gastrointestinal,
CC haemostatic, hepatotropic, immunostimulant, immunosuppressive, muscular,
CC nephrotropic, neuroprotective, osteopathic, respiratory, vasotropic and
CC virucide activities, and can be used in gene therapy. The nucleic acid

CC (I) and the encoded polypeptides, compositions, kits and methods are
CC useful in diagnosing and treating an immune related disease and in
CC stimulating an immune response. The present sequence represents a human
CC PRO protein from the present invention.
XX
XX Sequence 1972 AA;
QY
Query Match 6.7%; Score 532; DB 8; Length 1972;
Best Local Similarity 18.1%; Pred. No. 4e-26;
Matches 354; Conservative 256; Mismatches 599; Indels 750; Gaps 68;
QY 8 KPFPPLVPLGPEPF-FTIPH-----TQFAFRTRREYEARELRYSERIWTCKST 55
Db 189 KPLSLVNQAKETYMKLIVPSPDLKAGNKNTSESSLTSELRSREQV-----KQA 241
QY 56 GSSQLTKHEAWEEQEVAEELLK-----BEFPAYWEKLVLEMVHNN-----TA 97
Db 242 FFSQALKKQESSKSLKVIKVAALSNPKATSSPA-HPKQTLNHNPNFLTNALLGNHPNG 300
QY 98 SLEKLVDTAWLIMTYKAYGEECDPEV-----124
Db 301 VIQSVIQEAPLALTTKTKMQSKINENIAAASSTPSPVNLSTSGRRTPGNQTPVMPAS 360
QY 125 -----GKEKML--KVIIVKI---HPL-----EKVDEE 146
Db 361 PILHSQGEKAVSNVNVNPKTQHSHPAKSLVEQFRGTDSDIPSSKSDSEDSDEDEDE 420
QY 147 ATEKXSDGACDSPSDKENSQIAODHOKKETVVKVEDEGRRESINDRARRSPKLTSLK 206
Db 421 EDEDEDEDEDESDSDSSESDTEGSEBEDDDKQDQSDSD-----T 467
QY 207 KGERKWAPPKFLPHKYVDVKLQNEKIIINV--PADSLIRTERPPPKKEIVRYFIRHNALRA 264
Db 468 EGE-----KTSMKL---NKTTSVKSPSMLTCHGSTPRNLHIAK-----503
QY 265 GTGENAPWVVEDELVKKYSLPKSFDFLDLPYKYMTLNPSTKRNKNTGSDPRKPSKSK-T 323
Db 504 -----APGSAPALCSESQSPA-----FLGTSSSTL---TSSPHSGTSKERRVT 544
QY 324 DNSSLSPLNPKLWCHVHLKSLSGSPLKVYKNSKNSKSP-----EEHLEMMKMSNP---376
Db 545 DERELRIPLEYG-WORETRIRFGG---RLQGEVAYYAPCGKKLROYPEVIKYSRNGIM 600
QY 377 -----KLHT-NFHIPKKGP-----PAKKPG 395
Db 601 DISRDNFSFSAKIRVGDVFYEARDGQEMQWCLLKEDVIPRIRAMEGRGRPNPQRORA 660
QY 396 KHSKPLKAKGRSKGI-----LNGQKSTGNS 421
Db 661 REESRMRRRKGRPPNVGNVNAEFLDNADAKLLRLQAEIARQAAQIKLLRLKQKQEQARVA 720
QY 422 KSPKK--GLKTPKTKMKQMTLLDMAKGTQKMTAPRNSGGTPTSTSKPHKHLPPAALHLI 479
Db 721 KEAKQQAAMAAEKKRQKEQIKMKQKEKIRIQO-----IRMEKEL--RAQQL 769
QY 480 AYYKMKKREDKRSALSCVISTARLLSSEDDARLPEELRSLVQKRYELLEHKKRWASMS 539
Db 770 EAKKKKEE-----ANAKLLEAEKIK-EKENR---RQAVLLKHQER-----809
QY 540 EQRKEYLKKRBEELKKLKEKAKERREKEMLERLEKQRYBDEQL---TGKNV-----590
Db 810 ERRRHMMMLKMAEARKKAEERLQKEKRLNKRKEQRLLEMAKELKPKNED 869
QY 591 -----PAFLVDTPE-GLPNTLFGDVAMVVEELSCYSGLLLPDAQOYIPITAVSLMEALS 642
Db 870 MCLADQKPLPELPRIPGLVLSGSTFSDCLMVVQFURNFKGVDFVNDVPLNSVQEG 929
QY 643 ADKGFL-YLNRLVLIQTLQDE-IAEDY---GELGMKLSIEPLTLHSVSELVRLCLR 697
Db 930 LNIGDSMGEVQDLLVRLLSAAVCDPGLITGYKAKTALGEHLNVLGVNRDNLVSEILOIFM- 988
QY 698 RSDVQEESESGSDTDNKKOSAAFEDEVQDFLEKLETSEFFELTSEEKLOITALCHRL 757

Db 989 -----EAHCQGTETSLTKFAQHAHTPAQKASVLAFLINELA 1026
Qy 758 MTYSVQDMETRQQAELWKEKRLAVLKEENDKGAEKQKEMAKKENGKVGNG--- 814
Db 1027 CSKSVSEIDKNIDVMSNLRDKWV-----EGKLRKRIIHAKKTGKRDTSGLDLGEEQ 1082
Qy 815 --LG-----KTDRKRVKPEPQVDTAEADMISAVKSRRLAIAQAKKERIQE--REWK 864
Db 1083 HPLGPTPGKRRKRGSDYDDDDDDDD-----QGEDDEDEEDKEDQK 1129
Qy 865 VKLERQAEERIRKHAIAAEK-----AFQEGIAKAKLVMRRTIGTDNRNHR 911
Db 1130 GKKTIDCEDEGQAAASVEELEKQIEKLSKOOSQYRRKLPDASHSLASVMEGPDYRHR 1189
Qy 912 YWLFSDVEPLGPIE-----KGWH-----DSIDVRFNHC 941
Db 1190 YWIL-PRCGGIFVEGMEGEGLEETAKEREKLKKAESVQIKKEEMFETSGDSLNCSTNTHC 1248
Qy 942 -----KDHT---VSGDEDYCPRSK-----KANLGKNA----- 965
Db 1249 EQKEDLKEKDNINLFLQKPGFSKLSKLELVAKMPESEVMTPKPNAGANGCTLSYQNSG 1308
Qy 966 --SMNTOHGTATEAVE-----TTTTKQGGQLWFL- 993
Db 1309 KHSLSGVSTATQSNVERADSNLNTGSSGPKFYSPLPNDQLLKTILTEKNRQ--WESL 1366
Qy 994 -----CD----- 995
Db 1367 LPRTPCDDTSLTHADMSTASLVTPQSPKSPSPPTAPLAGSSAQNPVGLNPFALLSQV 1426
Qy 996 ----- 995
Db 1427 KGVSMGLQFCGWPVTGVTWSNIPPTLSVPSLGLSGLSGBNGNSFLSNVASSKSEPV 1486
Qy 996 -----SOKELDELLINCLHPQGIRES 1015
Db 1487 PQNEKATSAQPAAVEVAKVPDPPSPKPIPEEMQFCGWRIIDPEDLKALLKVLHLGIREK 1546
Qy 1016 QKEREKRYQDIHSHIHLAKPNGLKSCDGNQELLNPLASDLIE-----VATRLQKGGIG 1072
Db 1547 ALQKQIQK-HLDYITQACLKKN-KDVAITELNEE--NQVTRDIVENNVSBEQAMEMDLS 1602
Qy 1073 YVEETSEFEARVISLEKLDKFGECVIALQASVKKFLOGFWAPQO----- 1117
Db 1603 VLOQVEDLERRVAS-----ASLQ-----VKGWMCPEPASEREDLVYFEHKSP 1644
Qy 1118 -----KRRKQSEDSAKTEEYDEEKQWVE--EAKVASALEKWK 1153
Db 1645 TKLCKEHGFEFTGEDESSAHALERKSDNPLDIAVTRLADLERNIERRIEEDIAPGLRYVR 1704
Qy 1154 TAIREAQTFSRMHVLLGMLDACIKWMSAENARCKVCPKGBEDDKLILCDECNKAFHFLFC 1213
Db 1705 RALSEARSAQAVALCIIQLOKSIAMEKSIKMYCYQICRKGDNELLLCDGCKGCHTYC 1764
Qy 1214 LRPALYVDPGEWQCPACOPATA-----RNSRGRNYTESASE-----DSDDSD 1260
Db 1765 HRPKLTITPDGDMWFCPACIAKASGOTLKI KLUHVKGKNTNESKKGKKVTLTGDTDEDS- 1823
Qy 1261 EEEEEEEEBEEDYEVAGLRPRKTRIRGKHSVIPPAAARSRRRPGKPHSTRSQPKAP 1320
Db 1824 -----ASTSSSLKGNKDLQKRKEENTS 1847
Qy 1321 PVDDAEVDVLQTKRSSRROSLQKEEILHKIVKYRFSWPPREPVTREDAEYDVI 1380
Db 1848 --INLSKQESFTSVKPKRRDSDKDALCGLSMILTEMETHEDAWPFLLPVNLKLVPGYKKVI 1905
Qy 1381 THPMDFQTVQNKCSGYSRVSQEFELTDMKQVPTNAEVN 1419
Db 1906 KKPDMDFSTIREKSLSSGGQPNLETFAIDVRLVFDNCETFN 1944

AAW81170 standard; protein; 1878 AA.
AAW81170;
05-MAY-1999 (first entry)
Human BAZ2-alpha protein.
Transcriptional regulator; transcription; BAZ1-alpha; bromodomain; BAZ;
atypical zinc finger; testis; human; tumour; BAZ1-beta; BAZ2-alpha; drug;
BAZ2-beta; treatment; cancer; proliferative disorder; screening.
Homo sapiens.
WO9847920-A1.
29-OCT-1998.
17-APR-1998; 98WO-JP001783.
18-APR-1997; 97JP-00116570.
24-OCT-1997; 97JP-00310027.
(CHUS) CHUGAI RES INST MOLECULAR MEDICINE INC.
Jones MH;
WPI; 1998-583603/49.
N-PSDB; AAV68396.
Transcriptional regulator gene family containing bromodomain - may be
expressed in testis tissue and is useful for treatment of cancer and
other proliferative disorders.
Claim 1; Page 72-88; 187pp; Japanese.
This sequence represents the human BAZ2-alpha protein, a member of a
family of transcriptional regulator genes containing a bromodomain (BAZ,
Bromodomain with Atypical Zinc finger) which are expressed specifically
in testis tissue and also in certain tumour lines. Transgenic calls may
be used for the preparation of the BAZ1-alpha, BAZ1-beta, BAZ2-alpha and
BAZ2-beta proteins. These proteins can be used in the treatment of cancer
and other proliferative disorders, and in screening of compounds for
their binding ability to the expression products (e.g. for use as drugs
by modulation of transcriptional regulation)
Sequence 1878 AA;
Query Match 6.6%; Score 523.5; DB 2; Length 1878;
Best Local Similarity 19.4%; Pred. No. 1.4e-25;
Matches 265; Conservative 181; Mismatches 435; Indels 483; Gaps 43;
Qy 415 QKSTGNSKSPKGLKTPKTKOMTLLDMAKGTQ---KOTRAPRNSGGTPTSSKPHKL 471
Db 619 QAITGKRGRPR--NTEKAKTKEVPKVRGRGRPPKVI TEL-----LNKTDNRPLKCL 669
Qy 472 PPAALHLIAYYKENKDRKRSALSCVISTARLLSSEDRARLPBELSLVQKRYELLEH 531
Db 670 EA-----QETLNEEDKAKIAKSKKKMKQKQVORGECLTTIOGAARNKQKQETSKLKH 720
Qy 532 KKRWASMSSEQRKEVLLKKREELKKLKEKAKER---REKE----- 569
Db 721 KE--AKKSKAEKEKGKTKQEKLEKVKREKKEKVKMEKEEVTYKAKPACADKTLATQR 778
Qy 570 MLERLEKO-----KRYEQELTG-KNLPAFLRVLTDPFG--LPNTLFGDVAMVVEFLS 618
Db 779 RLEERQKQKQMILEEMKKPTEDMCLTDHQLPDPFSRV---PGLTLPAGAFSDCLTIVEFLH 835
Qy 619 CYSGLLLPDAQYPITAVS-LMEALSADKGGFLYLNRLVILLQTLQDEIAEDYGE---- 673
Db 836 SFGKVLGFDPAKDVPSLGLVQEGLLCOGDSLGEVQDLLVRLLLKAALHDPGFPFSYCSQLKI 895
Qy 674 LGMKLSLEIPLTLHVSSELVRLCLRRSDVQSESESGSDTDDNKDSAAAFEDNEVDQEFLEKLE 733

XX EP1033401-A2.
XX 06-SEP-2000.
XX 21-FEB-2000; 2000EP-00200610.
XX 26-FEB-1999; 99US-0122487P.
XX (GEST) GENSET.
XX Dumas Milne Edwards J, Duclert A, Giordano J;
XX WPI; 2000-500381/45.
XX N-PSDB; AAC03106.
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
XX obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
XX diagnostic, forensic, gene therapy and chromosome mapping procedures.
XX Claim 13; SEQ ID NO 7181; 71pp + Sequence Listing; English.
XX The present sequence is a polypeptide encoded by one of a large number of
XX 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs were
XX prepared from total human RNAs or polyA+ RNAs derived from 30 different
XX tissues. EST sequences usually correspond mainly to the 3' untranslated
XX region (UTR) of the mRNA because they are often obtained from oligo-dT
XX primed cDNA libraries. Such ESTs are not well suited for isolating cDNA
XX sequences derived from the 5' ends of mRNAs and even in those cases where
XX longer cDNA sequences have been obtained, the full 5' UTR is rarely
XX included. 5' ESTs are derived from mRNAs with intact 5' ends and can
XX therefore be used to obtain full length cDNAs and genomic DNAs. 5' ESTs
XX are also used in diagnostic, forensic, gene therapy and chromosome
XX mapping procedures. They are used to obtain upstream regulatory sequences
XX and to design expression and secretion vectors
XX Sequence 60 AA;
Query Match 4.4%; Score 352; DB 3; Length 60;
Best Local Similarity 100.0%; Pred. No. 3.8e-16;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 898 MRRTPGTDRNNHRYWLFSDVPLGFIKGVHDSIDYRNFHHCKDHTVSGDEDYCPRSK 957
DB 1 MRRTPGTDRNNHRYWLFSDVPLGFIKGVHDSIDYRNFHHCKDHTVSGDEDYCPRSK 60
RESULT 47
ABRS2663
ID ABRS2663 standard; protein; 1264 AA.
XX AC ABRS2663;
XX 20-JUN-2003 (first entry)
XX Protein sequence #SEQ ID 191.
XX Multiprotein complex; eukaryote; drug target; diagnosis.
XX Saccharomyces cerevisiae.
XX EP1258494-A1.
XX 20-NOV-2002.
XX 20-DEC-2001; 2001EP-00130253.
XX 15-MAY-2001; 2001EP-00111774.
XX (CELL-) CELLZONE AG.
XX Bauer A, Gavin A, Grandi P, Krause R, Kruse UD, Kuester BD;
PI Marzioch M, Schultz JD, Superti-Furga GD;

XX WPI; 2003-250078/25.
XX N-PSDB; ACC60705.
XX New isolated protein complexes useful for diagnosing a disease or
XX disorder, or as a target for an active agent of a pharmaceutical,
XX preferably a drug target in the treatment or prevention of disease or
XX disorder.
XX Disclosure; SEQ ID NO 191; 17pp + Sequence Listing; English.
XX The invention relates to multiprotein complexes from eukaryotes. Proteins
XX of the invention and DNA sequences encoding them are given in records
XX ABR52568-ABRS3903 and ACC60610-ACC61944 respectively. The complexes are
XX obtainable by using a protein as a bait and isolating the set of proteins
XX which is attached thereto from cells. Such protein complexes may comprise
XX up to 30 distinct proteins. Protein complexes of the invention are useful
XX for diagnosing a disease or disorder, or as a target for an active agent
XX of a pharmaceutical, preferably a drug target in the treatment or
XX prevention of a disease or disorder. Note: The sequence data for this
XX patent is not represented in the printed specification, but is based on
XX sequence information supplied by the European Patent Office. The complete
XX document is available on CD-ROM
XX Sequence 1264 AA;
Query Match 4.4%; Score 347.5; DB 6; Length 1264;
Best Local Similarity 19.5%; Pred. No. 5.4e-14;
Matches 271; Conservative 206; Mismatches 492; Indels 419; Gaps 58;
QY 3 PILGRRKPPFLVNPPLGCEBFFTPHTQEAFTREEEARLERYSERIWTCKGTGSQLTH 62
DB 8 PILLPDPKPL--PLDLNVQVWHIEETGEWSSVEEFLERDFYTRHFTCEITGTSLTF 65
QY 63 KEAWSEEQVAELLKEEPPAWYEKLVLEMVHN--TASLEKLVDTAWLEMTKYAVSECD 121
DB 66 FOALDSEETQFYKVEDDFPLKLRPEVARELHFGIRLDALVEKVFARPKNDFFPGE--- 122
QY 122 FEVGKEMKLVKIVKHPLEKVDDEATEKKSGDAGCDSPSSDKENSS---QIAQDQKKE 177
DB 123 -----VVYLRKQKDSSTSSNQSQSTPQDDMVVINSVGNPGLPYQYQRR 168
QY 178 TVVKEDEGRESINRARRSRPKLPTSLKKGKRWKAPKFLPHKYDVKLQNEDKIISN-- 235
DB 169 VYIKEKQVGNATINPESREIVMPAHT-----KY---MLIEAASNSKS 208
QY 236 --VPADSLIRTERPPNKEIVRYFIHNALRAGTGENAPWVDELVKYSLSKPSDFLL 293
DB 209 FIVDQGIYDRSTFTFKHLIKCFKITLQASSKMGAPCVRKPEYLYAMYGLTWEPK--- 265
QY 294 DPYKMTLNPSTKRK---NTGSPD-----RKPSKSKTDNSSLSSPLNPKLMCHVHL 342
DB 266 DMLKYKEDFPVARRSNSANVSSPESEKMKRSKSGKSNNTNDASNKETK--- 318
QY 343 KXSLSGSLPKVKNKSNKSPPEHLEEMMMKMSPNKLHTNFHIPKKGPPAKPGHSDKPL 402
DB 319 ----KRPKTEVNDSENNSEED-----KKKGQNVTS 347
QY 403 KAKGRSKGLNGCKSTGNSKSPKGLKTPKTKQKQMTLL-----DMAKTKQKTRAPRNS 457
DB 348 HSKKRKKE--ANEENPTENVES-----VPTPANAEPAQVTTITSDMDLA----- 389
QY 458 GGTPTRTSSKPHKHLPPAALHLIAYYKENDKREDKRSALSCVISTARLLSSEDRARLPBE 517
DB 390 -----LPYQH--PPNIFPNLTYYNEK-----LECI-----SLGSTKLSRPFDS 425
QY 518 LRSLVQKRYELLE---HKRWASMSSEEQKRYLK-----KREEL 554
DB 426 FGKLLQ-AYQFLNTFGSKICLSHFSLDQFITSCLKTDPVELKGEVVLVNVNIRTQTSKEQBI 484
QY 555 KKK---LKEKAXERKEKMERLEKQRYEDELGTGK---NLPAFLV-DTEG---LPN 604
DB 485 ENNGLPMKNKAETTTTEEDSENPSQWRNSFIRDMIMKRNDSKVYKIVHDDPASDILDN 544

Db 348 HSKRKKE-ANEENPTEENVES-----VPTANAEPPQAVTITSIMDDLA----- 389
Qy 458 GGTPTSSKPHGLPAALHLIAIYKKNKREDKRSALSKVISTARLLSSSDRAALPPE 517
Db 390 -----LPHYQH-PNIPFNLYNEK-----LECI-----SLGSTKLSRPPFS 425
Qy 518 LRSLVQRYELLE---HKRWASMSBEQRKEYLK-----KREEL 554
Db 426 FGKLLQ-AYQFLNTFGSKICLSHSLDQFTLSLKCTDPVELKGCVLVNIRQTQSKEQEI 484
Qy 555 KKK---LKEKAKERREKEMLERLEKQRYEDQLTKG---NLPAFLV-DTPEG---LPN 604
Db 485 ENNGLPKMKAEATTTSSEDSNPDSQWQNSFIRMIMKRNDSKVKEYKIVHDDPASDDILDN 544
Qy 605 TLFQGVAMVVEFLSCYSGILLPPAQYPIITAVSLMEALSADKGGFLYLNRLVILLQTLQ 664
Db 545 INHGSALLIEVFTALLRLFI-NEEGDWSCIVVENWIIDKG-----VLMER--K 591
Qy 665 DEIAEDYGLGKMLSEIPLTLHSVSELVRL--CLRR--SDVQSESEG-SDTDNKNDSAAF 719
Db 592 DERGE--GEAKQRNAHGYFLQDKEKIDNLKOTLKENATEVQKESDAKNETNSESDSKSD 649
Qy 720 EONEVODEFEK-----LETSEFEFELTSE----- 743
Db 650 SDEERDPKLEKCLNRYNNVNIERLTQROFNNSYMLIILGLVLEDCRHLPMYTEFIDSFI 709
Qy 744 EKL-----OILTALCHRI-----LMTY-----SVQDHMETRQ 770
Db 710 EKLIIPKDISATQPKQLWRNFCFKLSFKVNALWILVDLVSHFSFDPIKAAVDSDME--- 766
Qy 771 QMSAELWKERLAVLKEENDKRAEKOKRAMEA-----KNKENGKVENGLGKTRDKRI 824
Db 767 -LCGQIRSERFKVARELKTEAAVLSNLQDQLOAIQELKNKTIDENTPSADGDKD----- 820
Qy 825 VFEPQVDTEADMI SAVSRLLAIQAKEBIEIQRKMKVKLEQROAEERIRKHAAB 884
Db 821 ---DSESNSEPIDLIITIEKKQKLIIEQDKKVQALQ-----SD 854
Qy 885 KAFQEGIAKAKLVMRTPITGTDNRNRY--WLSFDEVP-----GLFIEKGWVHSDIDY-- 935
Db 855 KNFLDNCLFENDLQRLKPLGLDRYGNRYFWLDHNGVFPQYPQAGM-NETPKNNSLSYHS 913
Qy 936 -----RFNHCKDH-----TVSGDEDYCPRS 956
Db 914 GRLLIQGPASSAKPFLNVSDQLSNWQKIRNSEGISEATREVFGISKTSSGSYNYVENG 973
Qy 957 KKAN-LGKNASWNTQ-HGTATEVAVETTPKQ---GQNLWFLCDSQKQELDELNCLHPQG 1011
Db 974 IEVELLSDNDRVNLPIELTPIQKKIMDETFSRLLLSPDQWYCIDKLEDLSRIMDWLDNWG 1033
Qy 1012 IRESOLKERLEKRYODI IHSIHLARKPNLGLKSCDGNQELL-----NFLRSDLI 1060
Db 1034 RKEHLLRQIRIMERIKSLSL-RDHALSUFTAKNEEKLKLENNEFTENELNVDSM 1092
Qy 1061 EVATRLQKGLGVTESTSEFARVISLEKLKDFGECVIALQASVIKFFLQGFMAPKQKR 1120
Db 1093 DVDDK--NSGV-----KSEVDVQVDAEK---REAVIDEKLEVIADL----- 1130
Qy 1121 KLOSEDSATEEV-----DEEKQWEE-----AKVASALEKWKATAREAQ 1161
Db 1131 -MKLDDSSXTRNVNLRIQEBELQDRLQKQKSIINSQRPGARILARSERKTKISRGNK 1189
Qy 1162 FSRMHVLL 1169
Db 1190 VNKQTEIL 1197

RESULT 49

ABB58514

ID ABB58514 standard; protein; 2897 AA.

XX

AC ABB58514;

XX 26-MAR-2002 (first entry)
XX Drosophila melanogaster polypeptide SEQ ID NO 2334.
DE Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
KW Drosophila melanogaster.
OS WO200171042-A2.
XX 27-SEP-2001.
XX 23-MAR-2001; 2001WO-US009231.
XX 23-MAR-2000; 2000US-0191637P.
PR 11-JUL-2000; 2000US-00614150.
XX (PEKE) PE CORP NY.
XX Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-656860/75.
DR N-PSDB; ABL02617.
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signaling and cell-cell
PT interactions.
XX Disclosure; SEQ ID NO 2334; 21pp + Sequence Listing; English.
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 2897 AA;
Query Match 4.3%; Score 344; DB 4; Length 2897;
Best Local Similarity 19.1%; Pred. No. 3e-13;
Matches 265; Conservative 182; Mismatches 442; Indels 500; Gaps 55;
Qy 61 THKEAWEE---EQBVAELLKEEFPAPWVEKLVLEMVHHTASLEKLVDTAWLEIMTKYAV 116
Db 603 TESEALIEGVRNRTEVSVLQNNNSDSSSIQTQOHKVPADLNN-----TKKPV 652
Qy 117 GEECDFFVGVKEMKLVKVIKHPLEKVEDEATEKSGDSCDPSDSEKSSQIAQHQKK 176
Db 653 TEKCVLESG---TVEDKLVINH-----QQISTNALDTACDEKLS-CETESPVPNQHDSK 702
Qy 177 ETVVKED-EGRSEINDRARSPRKLPSTSLKGERKWAAPPKLPKHYDYKLVQNEKILSN 235
Db 703 MKSEQOTATSAKESVSESS-----PTD---GEDETAKSKCLIDPYAGKTNDKGIINK 752
Qy 236 V-PADSLIRT---EREPNKEIVRYFIRHNALRAGTGENAPWVVEDELVKVYSLPSKFSDFL 292
Db 753 IDKAEGVITLLODDPPDKDISKSEINTTCLINISPSAEHRIS----- 796
Qy 293 LDPYKYMTLNPSTKRAK-----NTGSPDRPKSKSKTDSNSSLSPNPKLVCHVLKKS 345
Db 797 ----KVQTEPSTSQNLLFEDNGKSGEVDGKSRSTGAVEEISKTSTL-----LNRKRR 845
Qy 346 LSGSPKLVKNSKNSKSPHEHLEEMKMSPNKLTNFIHPKKGPAKPKGKHSKDPKAK 405
Db 846 LNDQSALRNS--TSESEVHEEE-----PQDDPTLDDLDVGGKRIKMR 887

```
QY 406 GRSKGLNQKSTGNSKPKGLKTPKTKMKQMTLLDMAKGTQKMTAPRNSGGTPRTSS 465
Db 888 PK-----TTNAESRRK-VEAQKQIBETT-----SSGEDDARTRKCI 925
QY 466 KPHKLPAPAAHLAIYKENDKREDKRSALS CVISKTARLLSSDRALPEELRSLVQR 525
Db 926 AHTTKPTLEBII-----ERLKKSI-----EMDLPEKTEKVMEM 962
QY 526 YELLBHKRWA-----SWSBQRKYLEKKRRELKKLKEKAKERREKEMLERLEKQRY 580
Db 963 PTQLMQKTREATPPIISLSPOKKTPTTKPLKNLLTLQLRQ-----EESBETIPR-----KRT 1015
QY 581 EDQELTKNLPAFLVDVPEGL-----PNTLFGDVAMVVEFLSCVSGLLLPDAQYP 631
Db 1016 NSETLV-PAIPASNLVPCQDERHRRKRSSEDAEAFSKESPTE-----VP 1060
QY 632 ITAVSLMEALSADKGFYLNRLVLLQLTLQDEIAEDYGLGMKLSIPLTHSVSEL 691
Db 1061 PSAVS--EKLKENN-----EQDIQEV-----EDPLAM-SVKDS 1091
QY 692 VRLCLRRSDVQESE-----GSDTDNDKDSAAFEDNEVDQEFLEKLTSEFFE 739
Db 1092 LRSADQSPVPGSARRSGRRGPAVHSSELPOPKRTGGARDKQVPEVNAELKQ-----E 1147
QY 740 LTSEKQLQTLTALCHRLMTYSVDHMETRQMSAELMKERLAVLKEENDKKAQKQK 799
Db 1148 SEDDEKI-----STKIKSEAKODPAPE-----SPENRKKPEEKPTKE 1184
QY 800 EMEAKNKENGKVENGLGKTRKRVKPEQVDT-----EAEDMISAVKSRRLLAIOAKE 855
Db 1185 E-----PNEBPKVGRGPRKR-----EVDTTNIIETNDSETFVRSRRIAOQKKE 1234
QY 856 REIQEREMKVKLERQAEERIRKHAAKAFQEGIAKAKLVMRRTPTGTDNRHNRWLF 915
Db 1235 EAERKQOEVALRTMQE--LKKKKKAKEA----- 1263
QY 916 SDEVGLFIEKGVWVHSDIDYRNFHCKOHTV---SGDEDYCPRSKKANLGKNASNTOHG 972
Db 1264 -----DPTVLEPSGEE-----SES 1277
QY 973 TATEVAVETTPKQ-----GQNLWFLCDSOKELDELLNCLHPQGIRESOLKERLEKYODI 1028
Db 1278 EASEAEAEARNKKKCPKQGW-SSDSEEQ-----PSEEEEEEPHYE--- 1321
QY 1029 IHSIHLARKPNLGLKSCDGNQBLNPLRSLDLIEVATRLQKGLGYVETSEPEARVISLE 1088
Db 1322 -----TDPGSPL---FRSD-----HEFSPE-----S 1339
QY 1089 KLKDFGECVIALQASVKKFLOGFMAPKQKRKLQSEDSAKTVEEVEDEKKMYEAKVASA 1148
Db 1340 ELEDESQVV-----PMKARTVRKENADDDLEEDAE----- 1371
QY 1149 LEKWKTAIREAQTFSRMHVLLGLMDACIKWMSAENARCKVCPKKGEDDKLILCD--ECN 1206
Db 1372 -----ACQCKGSDHPWILLCTPTCN 1394
QY 1207 KAFHLFCPLPDIYVDPGQWCPAQ-----PATARRNSRGRNYTEESAS 1251
Db 1395 KGYHCSCLSPVFIYIPEGDWHCFPCQOEQLIAALERQLEQYDTLVAAQOERILAEQAE 1454
QY 1252 ESEDEDES-----DEEEEEEDEEEDYEVAGLRLRPRKTIRGKHSVIPPAARSG--R 1303
Db 1455 RERQELEAATLAAKOENFKSEKEEDDRDDMAVGKAEVKRRRGDGRINRRAAKRGTRR 1514
QY 1304 RPKKPHSTRR-----SQPKAPPVDDAEVDELVLQTKRSRRRQSLQLOK 1348
Db 1515 RGNESDSHRKSLGSGSGSGSSSDNSSTFSDDSDDEPIYKLRQINVSRLNEY 1574
QY 1349 BEILHKVIV 1357
Db 1575 DDLINSALK 1583
```

```
RESULT 50
AAM14835
ID AAM14835 standard; protein; 65 AA.
XX
AC AAM14835;
DT 12-OCT-2001 (first entry)
XX
DE Peptide #1269 encoded by probe for measuring cervical gene expression.
XX
KW Probe; human; microarray; gene expression; cervical epithelial cell;
KW cervical cancer.
XX
OS Homo sapiens.
XX
PN WO200157278-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US0000670.
XX
PR 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX
(MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-488901/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for analyzing
XX gene expression in human cervical epithelial cells.
XX
PS Claim 27; SEQ ID NO 19661; 487pp; English.
XX
CC The present invention relates to human single exon nucleic acid probes
CC (SENP; see AAI10068-AA128459). The present sequence is a peptide encoded
CC by one such probe. The SENPs are derived from human HeLa cells. The SENPs
CC can be used to produce a single exon microarray, which can be used for
CC measuring human gene expression in a sample derived from human cervical
CC epithelial cells. By measuring gene expression, the probes are therefore
CC useful in grading and/or staging of diseases of the cervix, notably
CC cervical cancer. Note: The sequence data for this patent did not form
CC part of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 65 AA;
Query Match 4.3%; Score 341; DB 4; Length 65;
Best Local Similarity 100.0%; Pred. No. 2.3e-15;
Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 233 ISNVPADSLITERPPNKEIVRYFIRHNALRAGTGENAPWVVEDELVKYSLPSKFSDFL 292
Db 1 ISNVPADSLITERPPNKEIVRYFIRHNALRAGTGENAPWVVEDELVKYSLPSKFSDFL 60
QY 293 LDPYK 297
Db 61 LDPYK 65
Search completed: March 9, 2005, 14:06:31
Job time : 151.673 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 9, 2005, 13:38:21 ; Search time 30.149 Seconds
(without alignments)
4885.990 Million cell updates/sec

Title: US-10-702-148-29
Perfect score: 7985
Sequence: 1 MAPLLGRKPPFLVNLPGEE.....NCFMMLVNTQFCMALTDVT 1531

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 65 summaries

Database : PIR 79:*
1: piri:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6977	87.4	1479	2 T17401	transcription regulator
2	569	7.1	811	2 T08738	hypothetical prote
3	459.5	5.8	1378	2 G88637	protein F53H1.4 li
4	450	5.6	1430	2 T34516	hypothetical prote
5	388	4.9	449	2 T12495	hypothetical prote
6	345.5	4.3	1264	2 S64146	probable membrane
7	325.5	4.1	3488	2 T34418	hypothetical prote
8	322	4.0	1790	2 S67593	transport protein
9	310.5	3.9	1526	2 A45605	mature-parasite-in
10	302.5	3.8	3225	2 I52300	giantin - human
11	302.5	3.8	3259	1 A56539	giantin - human
12	298.5	3.7	1372	2 JC5420	smooth muscle myos
13	296.5	3.7	2116	2 A25655	myosin heavy chain
14	295.5	3.7	3187	2 JC5837	364K Golgi complex
15	294	3.7	1909	2 A45592	liver stage antige
16	290	3.6	1827	2 T16270	hypothetical prote
17	288.5	3.6	5327	2 T13564	microtubule-associ
18	287.5	3.6	1938	2 JC5421	smooth muscle myos
19	286.5	3.6	2139	2 T18296	myosin heavy chain
20	285	3.6	4687	1 A39638	plectin - rat
21	283	3.5	2057	2 S61477	myosin II heavy ch
22	281.5	3.5	1979	1 S03166	myosin heavy chain
23	280	3.5	2017	1 A36014	myosin heavy chain
24	279.5	3.5	1972	1 A41604	myosin heavy chain
25	279.5	3.5	2663	1 S28261	centromere protein
26	279	3.5	1819	2 A71928	cag island protein
27	278.5	3.5	1390	2 S51364	sperm tail-specifi
28	275.5	3.5	2020	2 T21174	hypothetical prote
29	275	3.4	1927	2 G64585	cag pathogenicity

30	274	3.4	4684	2 A59404	plectin [imported]
31	273	3.4	1410	1 A57013	early endosome ant
32	272.5	3.4	3924	2 S37431	ankyrin 2, neurona
33	272	3.4	1898	1 A45973	trichohyalin - hum
34	265.5	3.3	1432	2 B85431	trichohyalin like
35	265.5	3.3	1871	2 D96796	probable heat shoc
36	263.5	3.3	1474	2 T20488	hypothetical prote
37	263.5	3.3	1964	2 A52882	nonmuscle myosin I
38	263	3.3	1976	2 A52852	myosin heavy chain
39	262.5	3.3	1938	1 A40997	myosin heavy chain
40	262.5	3.3	1992	2 A47297	myosin heavy chain
41	261.5	3.3	678	2 A54514	glutamic acid-rich
42	261	3.3	853	2 T51505	hypothetical prote
43	261	3.3	2007	1 B43402	myosin heavy chain
44	261	3.3	4574	2 G02520	plectin - human
45	260	3.3	1805	2 A34736	plectin - rat
46	260	3.3	2442	2 T08621	centrosome associa
47	259.5	3.2	1939	2 T18372	repeat organellar
48	259.5	3.2	1999	1 S21801	myosin heavy chain
49	258.5	3.2	2068	2 A47371	transcription init
50	257.5	3.2	1934	2 I48153	myosin heavy chain
51	257	3.2	1110	2 I51116	NF-180 - sea lamp
52	257	3.2	1638	2 A42091	transcription acti
53	255.5	3.2	1837	2 T41023	probable nuclear p
54	255.5	3.2	2253	2 T30336	nuclear/mitotic ap
55	255.5	3.2	2748	2 S57976	nuclear migration
56	255	3.2	1392	2 A43336	microtubule-vesicl
57	255	3.2	1572	2 S45251	SNP2alpha protein
58	255	3.2	1818	1 S73852	hypothetical prote
59	254	3.2	1957	2 T38077	hypothetical coile
60	254	3.2	2385	2 A32491	myosin heavy chain
61	254	3.2	2411	2 B32491	myosin heavy chain
62	253.5	3.2	2094	2 S33124	tpa protein - huma
63	252.5	3.2	1427	2 S22695	reslin - human
64	250.5	3.1	2954	2 T14156	kinasin-related pr
65	250	3.1	1939	2 I48175	myosin heavy chain

ALIGNMENTS

RESULT 1

T17401

transcription regulator WBSR9 - mouse

N:Alternate names: Williams-Beuren syndrome deletion transcript 9 homolog

C:Species: Mus musculus (house mouse)

C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004

R:Accession: T17401

R:Peoples, R.J.; Cisco, M.J.; Kaplan, P.; Francke, U.

Cytogenet. Cell Genet. 82, 238-246, 1998

A:Title: Identification of the WBSR9 gene encoding a novel transcriptional regulator,

A:Reference number: Z18735; MUID:9907764; PMID:9858827

A:Accession: T17401

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1479 <PRO>

A:Cross-references: UNIPROT:Q92277; EMBL:AF084480; NID:g4165089; PID:g4165089; PIDN:AAD

C:Genetics:

A:Gene: Wbsr9

A:Map position: 5

F:1360-1415/Domain: bromodomain homology <BRO>

Query Match 87.4%; Score 6977; DB 2; Length 1479;

Best Local Similarity 90.0%; Pred. No. 3.1e-305; Matches 1343; Conservative 67; Mismatches 66; Indels 16; Gaps 8;

QY 1 MAPLLGRKPPFLVNLPGEEPTTPTHTQEAFTREEYEARLERYSERIWTCKTSSQL 60

DB 1 MAPLLGRKPPFLVNLPGEEPTTPTHTQEAFTREEYEARLERYSERIWTCKTSSQL 60

QY 61 THKEAWEEQVAAELKKEFFPAWYKLVLEMHVHTASLEKLVDTAWLWIMTKYAVGEC 120

DB 61 THKEAWEEQVAAELKKEFFPAWYKLVLEMHVHTASLEKLVDTAWLWIMTKYAVGEC 120

```
QY 121 DFEVGKEMLVKVIKIHLEKVDREATEKKSDGCDSPSSDKENSSQIAODHQKETVV 180
Db 121 DFEVGKEMLVKVIKIHLEKVDREAVEKKSDGACDSPSSDKENSSQIAODLQKKEVV 180
QY 181 KEDEGRRESINDRARRSPKLPSTSLKKGKRWAPPKFLPHKYDVVKLQNEDKIISNPADS 240
Db 181 KEDEGRRESINDRARRSPKLPSTSLKGERKWAPPKFLPHKYDVVKLQNEDKIISNPADS 240
QY 241 LIRTERPPNKEIVRFIRHNALRAGTGENAPVVEDELVKKYSLPSPKSFDFLLDPYKWT 300
Db 241 LIRTERPPNKEILRFIRHNALRAGTGENAPVVEDELVKKYSLPSPKSFDFLLDPYKWT 300
QY 301 LNPSTKRNKGTSPDRPKSKSKTDSNLSSPLNPKLWCHVHLKSLSGSPLKVKNSKSK 360
Db 301 LNPSTKRNKGTSPDRPKSKSKPRDSSSLSPNPKLWCHVHLKSLSGSPLKVKNSKSK 360
QY 361 SPEEHLEEMKMWSP--NKLHNFHHPKGGPPAKPKGHSKPLKAKGRSKGILNGQKST 418
Db 361 SPEEHLEGVKMTSPNNKHLH-SFHHPKGPPAKPKGHSKPLKAKGRSGKILNGQKST 419
QY 419 GNSKSPKGLTKPTKTKMOMTLLDWAQGTQKMTAPRNSGGTPTSSPKPHKHLPPAALHL 478
Db 420 GNSKSPSKCVTKPTKTKMOMTLLDWAQGTQKMTPTPRSGGVPKSGPKPHKHLPPAALHL 479
QY 479 IAYYKENDREDKRSALSCVISTARLLSSDRAPLPEELRSLVOKRYELLEHKRWASM 538
Db 480 IAYYKENDKEDKKSALSCVISTARLLSNEDRAPLPEELRALVOKRYELLEHKRWASM 539
QY 539 SEQQRKEYLKKRKEELKKLKEKAKERKEKMLERLEKQRYEDQBLTGKNLPAFLVDT 598
Db 540 SEQQRKEYLKKRQELKELREKAKEREREMLERLEKQRFEDQELGGRNLPAFLVDT 599
QY 599 PEGLPNTLFGDVAMVVEFLSCYSGLLLPDAQYPIITAVSLMEALSADKGGFLYLNRLVIL 658
Db 600 PEGLPNTLFGDVALVVEFLSCYSGLLLPDAQYPIITAVSLMEALSADKGGFLYLNRLVIL 658
QY 659 LQTLQTLQDRIADYGLGKMLSEIPLTHSVSELVRLCLRRSDVQESGSDTDDNK 718
Db 659 ---LQTLQDRIADYGLGKMLSEIPLTHSVSELVRLCLRRCDVQESGSDTDDNK 715
QY 719 DSAAEPEDNEVDQEFLEKLETSFBLTSEEKLIITLALCHRLTMYTSVQDHWETQQMSA 778
Db 716 DSTPEDNEVDQEFLEKLETSFBLTSEEKLIITLALCHRLTMYTSVQDHWETQQVSA 775
QY 779 ELWKERLAVLKEENDKGRAEKQKREMEAKNKENGKVENGLKTDKRRIVKFPQVDTPE 838
Db 776 ELWKERLAVLKEENDKGRAEKQKREMEARNKENGKENVLGKVDKKEIVKIQQVEVE 835
QY 839 AEDMISAVKSRLLAIQAKEREIOEREMKVKLERQAEERIRKHAAAEKAFQEGIAKA 898
Db 836 ADDMISAVKSRLLSMQAKKEIEIOERETKVKLEREAEEERMKHAAAEKAFQEGIAKA 895
QY 899 KLVMRTPTIGTDNRNHRNRYLWFSDYVPGFLFIEKGWVHDSIDYRFNHHCKDHTVSGDEDYCP 958
Db 896 KLVLRTPTIGTDNRNHRNRYLWFNSVPGFLFIEKGWVHNSIDYRFKHRKDHSLNLPDDYCP 955
QY 959 RSKKANLGNASMTQHGATVAVETPTPKQGNLWFLCDSQKELDELLNCLHPQGIRES 1018
Db 956 RKKKANLGNASVNAHHGPALP-AVETTPKQGNLWFLCDSQKELDELLSCLHPQGIRES 1014
QY 1019 SOLKERLEKRYQDIITHSLHAKPNLGLKSCDGNQELLNFLRSLDIEVATRLQKGLGV 1078
Db 1015 SOFKERLEKRYQEIITHCNMARPNLGLKSCDGNQELLNFLRSLDIEVATRLQKGLGV 1074
QY 1079 ERTSFEARVISLEKLDGEGCVIALQASVIKKFLQGFMAPKQKRLQSDSAAKTEEVD 1138
Db 1075 EGTSTFEARVISLEKLDGEGCVIALQASVIKKFLQGFMAPKQKRLQSDSSTKSEVD 1134
QY 1139 BEKKVVEEAKVASALEKWKTAIREAQTFSRMHVLLGMLDACIKWDMSAENARCKVCPKKG 1198
Db 1135 BEKKVVEEAKVASALEKWKTAIREAQTFSRMHVLLGMLDACIKWDMSAENARCKVCRKG 1194
```

```
QY 1199 EDDKLILCDECNKAPHLCFLRPALYEVDPDGEWQCPACOPATARRNSRGRNRYTEESASEDS 1258
Db 1195 EDDKLILCDECNKAPHLCFLRPALYEVDPDGEWQCPACOPPTARRNSRGRNRYTEESTSEGS 1254
QY 1259 EDDSES-DEEBEEEEEEEEEDYEVAGLRPRKTIIRGKHSVIPPAARSGRRPGKXPHSTR 1317
Db 1255 EGDSEGEEBEEEEEEEEEDYEVAGLRPRKTIIRGKQSVI-PAARPGRPPGKXSHPAR 1313
QY 1318 RSQKAPVDDAEVDDELVLQTKRSSRRSLELQKCEBILHKLIVKYRFSWPPREPVTTRDEA 1377
Db 1314 RSRPK----DDPEVDLVLQTKRISRRQSLQKCEDILHKLIVKYRFSWPPREPVTTRDEA 1369
QY 1378 EDYDVITHTPMDFTQVQNKSCGYSYRQVEFLTDMKQVFTNAEVNCGSHVLSQWVKTE 1437
Db 1370 EDYDVIEHPMDFTQIYNKSCGNYRSQVEFLTDMKQVFANALNCGSHVLSQWVKTE 1429
QY 1438 QCLVLLHKLHFGHPYVRRKPKFPDLAEDEGSDSEPEAVGQSRDEDRRSRE 1489
Db 1430 QCLLALLQKHLPGHPYVRRKRPDLADDEGSDSESQSR--GRRQKK 1479

RESULT 2
T08738
hypoetical protein DKFp586E0518.1 - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 09-Jul-2004
R:Accession: T08738
R:Wambutt, R.; Heubner, D.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, March 1999
A:Reference number: Z16471
A:Accession: T08738
A:Molecule type: mRNA
A:Residues: 1-811 <WAM>
A:Cross-references: UNIPROT:Q9NRL2; EMBL:AL050089
A:Experimental source: adult uterus; clone DKFp586E0518
C:Genetics:
A:Note: DKFp586E0518.1
F:709-764/Domain: bromodomain homology <BRO>

Query Match 7.1%; Score 569; DB 2; Length 811;
Best Local Similarity 24.4%; Pred. No. 2.7e-18;
Matches 184; Conservative 115; Mismatches 239; Indels 216; Gaps 24;

QY 876 BEERIRKHAAAEKAFQEGIAKAKLVMRTPITGTDNRNHRNRYLWFSDEVPGLFIEK--GW 932
Db 29 BEALKQHQKKELEKIQSAIACTNIFPLGRDRMYRRYWF-PSIPGLFIEDYDGL 87
QY 933 VHDSEI----DYRFNHHCKDHTVSGDEDYCPRSKKN-LGKNASMTQHGATVAVETT 986
Db 88 TEDMLLPSPSFQNNVQSDPQVS-----TKTGEPLMSESTNIDOG-PRDHSVQLP 138
QY 987 TPQKQNLWFLCDSQKELDELLNCLHPQGIRESQKRL--EK-----RYQDIHISI 1036
Db 139 KPVHKPNRWCYSSCEQLDLTEALNSRGHRESALKETLQEKSRICQLARFSE--EKF 196
QY 1037 HLAKPNLGLK-----SCDGNQ-----ELLNFRSLDIEVATRLQKGLGVVET 1081
Db 197 HFSKDPQDPKPTYSRGRSSNAYDPSQCAKQLELRDLRDDIEDRIYQGTLAGIKVT 256
QY 1082 -----SEFEA---RVISLEKL-----KDFGECVIALQASVIKKFLQGFMAPKQKR 1123
Db 257 DBHIWRSALSGRYELLSEENKENGIIKTVNEDVEEMEIDETQKIVK--DRLLGIKTET 314
QY 1124 RKLOSDSAAKTEEVD-----EKQWEEAKVAS-----ALSKWKA 1159
Db 315 PSTVSTNASTPQSVSSVVHYLAMALFQIEQGIERRFLKAPLADSDSGRSYKTVLDRWRES 374
QY 1160 IREACTFSRMHVLLGMLDACIKWDMSAENARCKVCPKKGEDDKLILCDECNKAFHLCFL 1219
Db 375 LLSSASLSQVFLHSLTLDRSVIWSKSIILNARCKI CRKKGDAENMWVLCDCGRGHHTYCVR 434
QY 1220 PALYEVDPDGEWQCPACOPATARRNSRGRNRYTEESASEDSDEDEDEE-----EEEEEE 1275
```


Db 435 PKLKTVPEDWFCPCRCRQKQRRRLSSRRORPSLESDVEDSMGDEDDVEDDDEEGQSE 494
QY 1276 EBDYEV-----AGLRLRPKTI-----RGKSHVIPPAAARSRRRPGKPKPHSTR 1317
Db 495 EBEYEVEDDDQSEEEVSLPKRGPQVRLPKTRGLSSSFSSRGQQQEGRYPSSQ 554
QY 1318 RSQPK-----APPVD-----DAVDEL----- 1334
Db 555 QSTPKTTSVTSKTRSLRKINSAPPTETKSLRIASRSTRSHGFLQADVVELLSPRRKR 614
QY 1335 -----VLQTKSSRQSL----- 1347
Db 615 GRKSANKTPSPNPNFVIATKSEQRSSQSNIAISKLSQSESKRRCKRQSPSPV 674
QY 1348 -----BLQKCEELHKIVKRFSPFPREPVTVRDEAEDYDVIVTHPMDFTQV 1393
Db 675 TLGRSSRGQGVHLSAEQLVVELVRHDDSWPFLKLVSKIQVDPYDIKKPIALNII 734
QY 1394 QNKSCGYSRVOEFLTDMKQVFTNAEVNCRGS 1427
Db 735 REKNKCEYKLASEPTIDIELMFSNCFYNPRNT 768

RESULT 3
G88637
Protein F53H1.4 [imported] - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 09-Jul-2004
C:Accession: G88637
R:Anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biology
A:Reference number: A75000; WUID:99069613; PMID:9851916
A:Note: see websites genome.wustl.edu/gsc/C.elegans/ and www.sanger.ac.uk/Projects/C.ele
A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
A:Accession: G88637
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1378 <STO>
A:Cross-references: UNIPROT:O45075; GB:chr_IV; PIDN:AAC02578.1; PID:G2854159; GSPDB:GN00
A:Gene: F53H1.4
A:Map position: 4

Query Match 5.8%; Score 459.5; DB 2; Length 1378;
Best Local Similarity 19.9%; Pred. No. 4e-13;
Matches 319; Conservative 194; Mismatches 531; Indels 557; Gaps 59;

QY 71 EVAELLKEFPAYWEKLVEMVHHNTASLEKLVDTAWLEIMTKYAVGEECDPEVGRKML 130
Db 74 ETEDLIDEKFFAG-EKLMSSGKEYTVVSSEKRGGLTYTMDGTGKIGHR---DLRRKGL 129
QY 131 KYKIVK---IHPLEKVDTEATEKSDGACDSP-----SSDKENSSQIAQ----- 171
Db 130 SVEBIKTAIEDAEFYDERQKVRBELLAENPIREIKKYAPIFSANRKSPTKTAQLSVA 189
QY 172 ---DHQKETVVEDEGRRESINDRARRSPKLPSTLKGKRWAPPKFPHKYDVVKLQNE 229
Db 190 EADSDVQVSVMDASGAAPKANSAMKTPRGAP-----RASGGPVILSSR---RLQEK 239
QY 230 DKIIISNVPADSLIRTPPNKBEIVRYFIRHNALRAGTGNAPVWVEDELVKYLSPKFS 289
Db 240 QKEKEKEKELEKKQKEQ-----BEKKQKEEEKAKK-LKEK-E 277
QY 290 DFLDPYKYMTLNPSTKRKNTGSPDRKPSKSKTQNSLSPLNPKLWCHVHLKSLSGS 349
Db 278 EXLKEKEKAARKAEKKEKNGTMDK-----FLKXD-TGS 311
QY 350 PLKVKNSKN-----SKSPEEHEEMKQMSPNKLHTNPHIPKGGPPAKKPGKHSKPLK 403
Db 312 P-----SSKQAPLPSFSKWEKRIAGVKKQ-----EDAWKRDLLEY 349
QY 404 AKGRS----KGILNGQKSTGNSKSPKGLKTPKTKMQMTLLDMAGTKQMTAPRNSGGT 460

Db 350 NEACSWCEKNSGNORST-----FENPIFKFSVQKLVDKAK----- 385
QY 461 PRTSKPHKPLPAALHLYAYKENKREDKRSALSVCISKARTARLLSSSDRLRPEELRS 520
Db 386 -----DR----- 388
QY 521 LVQKRYELLEHK--RWASMSSEQRKEYLKKRBEKKKLEKAKERREKEMLEKOK 578
Db 389 -----HMKGMKWAQAE--FKAEMSEKRELYQKFEPIKAWFNEDI----- 428
QY 579 RYDEOELTKNLPAP--RLVDTPEGLPNTLPGDVAMVVEFLSCYSGLLLPDAQYPITAVS 636
Db 429 ALDDLLVTGCEJLLELDKRVNCEBLLKCL-----EISQFFVSMRKILLWNE--ITABQ 481
QY 637 LMEALSADKGGPLYLNRVLVILLOTLLOLQDEAEDYGLGMLKSEIPLTIHVSSELV 696
Db 482 LDDLL--HGGFDGFRSTYKMIANLLETAQEKHEKAAHCNARLSEFPINEHTISELI 538
QY 697 RLCLRRS-----DVQEESEGS-----TDNKOSSAAFE 724
Db 539 RAFFIGTSTSPKRDGKRAARGAHDDDEEDGDEEMDSREKVVETEEKPEENVENGAAE 598
QY 725 DNEVODEFLEKLETSFELTSEEKLOILTALCHRLIMTYSVQDHMETRQOQSAELWKE-- 783
Db 599 NGEIGDDEENSES-----ELQKQILALFADSCHYELPAGAQ--LEVLCA-----KEV 647
QY 784 -----RLAVLKEENDKKRAEKQK-----RKMEAKNKE-----NGKVENGGLGKTDRK 825
Db 648 VHGLPIIREWFLRDANSEKLTETHEKANRIRNEMEQFHQOLQDFPIPIETESMTQTQR 707
QY 826 -----KRIVFEPQVDTAEADMI SAVKSRRLLAIOAKERIQEREMKVKLERQABEER 879
Db 708 EAEQSLRRREKLENQD-----LKELEENREATAREVD-DLER----- 746
QY 880 IRKHAAAEKAFQEGIAKAKLVWRPTPTGTDNRHRYWLF---SDEVP-----GLFIEKG 931
Db 747 -----IFRVYHIGNDRHLRKYWFAYSSDAAIIVQDFGTTSYEK 785
QY 932 VYHDSIDYRFNHHKDHVTSGDEDCPRSKKANLGKNSMNTQHGTAATEVAVETTPKQG 991
Db 786 WVRDCEKGF-----DVESSDVENRPEYEDLP-----TSSQS 819
QY 992 QNLWFLCDSQKELDELNLCLHPQGIRESQLEKREKRYQDIITHSHLARKPNLGLKSCDG 1051
Db 820 SETWYKLDTEPAIRQLMTQKTNGKREKLLKLYLRNMDDIISILRKEKQ---KKDDG 876
QY 1052 NOE-----L 1055
Db 877 EEEDEEASEEEDASAAAEENGEBKMETEQNGTGKFGFPETAAAAAABTQBAEKVARFTGR 936
QY 1056 LNFLRSDLIEVATRLQKGLGVVEETSBPEARVISLEKLDKDF-----ECVI- 1102
Db 937 FGLSKRTSELNDWKQSGISKIVDSQVFEARMLEANTLDEMKLVELTSLTIPVSCVIE 996
QY 1103 ---ALQASVIKFLQFGMAPKQKRKLQSD-----SAKT-----EYDEEK 1141
Db 997 KFPQNVAIATKFLNFSTILLEKWKVNFSDFFRAKCRIFNYFLPLPSHKTSKFLKLNQTM 1056
QY 1142 KQVEBAKVASALEKWKTA-----IREAQTSRHHVLLGLMDACI 1180
Db 1057 KITAGFGICAFIQKPKDVRLLPKKKECFSHLMRPRFCRVQBSNASCLHMLLAYDARI 1116
QY 1181 KWDMAENARCVCPKKGEDDKLILDCENKAPHLFCLRP---ALYE---VPDGSWQCP 1233
Db 1117 DOORTLPBELSCVCRKKTGTERKLMCKQCSTVPHYGCHRTISRALLFEPEGKEG-WWCA 1175
QY 1234 ACQATARNRNSGRNYTEESASESD-----DESDEEBEEBEEDYEVAGL 1284
Db 1176 KCTKEDRRR-----QLSEAKEDLRQKEGGDEEDHHGSGSDDEDEEDIVEET 1225
QY 1285 RLRPKRTIRKGHSVIPPAAARSRRRPGKPKHSTRSQPKAPP-----VDDAEVDELVLQT 1338

Db 1226 RGRSAKR-----KANAAMRDVLEFEGVLRTQTPAPPKPKQKTVVVPVLELFSNI 1275
 QY 1339 KRSSRRQSLQKCEIHLKIVKVRFSWPFREPVRTRDEADYVDVITHMPDQTVQNKCS 1398
 Db 1276 ERANPLRYKTLQOI-----PGQSRSTRNAQHENSRLPDIEQDLIDV----- 1315
 QY 1399 CGSYBSYQEFLLDMKQVFTNA---EVVNCRCGSHVLSWVK 1435
 Db 1316 ---YTSASQHLHLSQFFRHARGYIETHNPKRFSKFPFK 1353
 RESULT 4
 T34516
 hypothetical protein ZK783.4 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 09-Jul-2004
 C:Accession: T34516
 R:Favella, A.; Vaudin, M.
 submitted to the EMBL Data Library, August 1994
 A:Description: The sequence of C. elegans cosmid ZK783.
 A:Reference number: Z21536
 A:Accession: T34516
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1430 <FAV>
 A:Cross-references: UNIPROT:Q23590; EMBL:U13646; PIDN:AAC24421.1; GSPDB:GN000021; CESP:ZK
 A:Experimental source: strain Bristol N2; clone ZK783
 C:Genetics:
 A:Gene: CESP:ZK783.4
 A:Map position: 3
 A:Introns: 248/3; 373/3; 547/1; 593/2; 905/2; 1042/3; 1116/3; 1317/2; 1376/1
 F:1284-1339/Domain: bromodomain homology <BRO>
 Query Match 5.6%; Score 450; DB 2; Length 1430;
 Best Local Similarity 18.8%; Pred. No. 1.1e-12;
 Matches 306; Conservative 255; Mismatches 583; Indels 486; Gaps 61;
 QY 126 KEKMLKVIKVIHLEKVEATEKKSQDAGDS-----PSSDKENSQ----- 168
 Db 21 QOQLQOQLAKI---QKATASPSTKSTNGTSASTSAVSTSGTSSQNEAAQLQNLAKQ 77
 QY 169 -----IAQDHOKKETV-----KEDEGRRESINDARRSRPKLPTSLKKGERK 211
 Db 78 QIQQLAQRGALMAAQKQKQKAAADKAKEKEKQKAAKAAAAKASASTSSASAIPG 137
 QY 212 WAPPKFLPHKYVDVQLQNEDKIISNPADSLIRTPPNKEIVRYFIRHNALRAGTGENAP 271
 Db 138 LSPLELAQAQIAQIOW-----ALQQMWMTPOKQSM----- 167
 QY 272 WVEDELVKYSLPSKFSDFLLDPYKMYTLN--PSTKRNTGSPDRKPSKSKTD----- 324
 Db 168 -----BEATKK-----MMDMAKXKPGAVASTSSASTSSSTSSASTSSANNAA 213
 QY 325 NSSLSPLNPKLWCHVHLK-----KSLSGSPLKVNKSNKSKSP 362
 Db 214 NNAASNMMNVMWQILVAQMOQKQOQKQDQKADQAKAKELAKQOQKQEDVANKQO- 272
 QY 363 BEHLEEMMK-MMSPNKLHTNFHIPKGPAPKPGXHSKPLKAKGRSKGILNGKSTGNS 421
 Db 273 ---EELKFLMAQHOLN-----HOKGHEKQADAAALAAKVAALAAHRALES 315
 QY 422 KSPKGLTKPTKMKQMTLLDWAQTKQKT-----RAPRNSGGTPTPTSS 465
 Db 316 DSPEEGKKTNEAMLR-----LPIQLQWRQTCVRSIASAGVGDVSYFAP--CGKKLSTYS 369
 QY 466 KPHKLPALALHLIAYKENKDRKRSALSCVISKTARLLSSDR---ARLPEELSLIV 522
 Db 370 EVVRVLTNKSIIYIT--RDNLFNTKLIVGIFVFKQTEADETQOEREFAMFTEDDINKE 427
 QY 523 QKRYELLEH-KRWASMSBQKRYLKKK-----BELKKKQKJE----- 560
 Db 428 LTRLNVLFVFPKIQASTNSGVHEDDIKMSKIEPDEPLDPSLNDFTTELHVSQIMSG 487

QY 561 ---KAKERREKEMLERLEKQRYEDQELTGKNLPAFLVDTPEGLPNTLFGDVAMVVEF 616
 Db 488 VDEKIRERADDLLVINDV-----RHLPDFSRIGN-QCLSQSGFADALMVHBF 536
 QY 617 LSCSYGLLLPDAQYPITAVSLMEALSADGGFLYLNRVLVILLOTLLQTLQTLQDEIAEDYG 676
 Db 537 VQNFHVLGIDLEIAPKLESACAGLDGDAN--HAEQTLQLTRQ--LLRLALEFPGMGNEK 592
 QY 677 ELGKMLSEIPLTLHSVSELVRLCLRRSDVQSESGSDTDNDKDSAAFEDNEVDQFLEKL 736
 Db 593 RFGQGGGEMGLDRENFSEVMRLFL-----IDKGRG-----EELSQPL 630
 QY 737 ETSFEFFELTSEKQILTALCHRIILMTYSV---QDHMETRQMSAEIWM----- 781
 Db 631 LTCNPLSISPQKASILAFLCDELVCSSRNVTETDKNDEISRLKGEKMRGKARALRS 690
 QY 782 -----KERLAVLKEE--NDKKRAEKQKRKE-----MEAKNKENGKVE 816
 Db 691 ARSKKNDKVVVVKEQNHESDSEPTPTPDTPKKATVAPPTVSVSPVSAQQQQRKFT 750
 QY 817 NGLKTRDRKRIKVFEPQVDTAEADMI SAVKSRRLALAIQAKKERIQEREMKVKLERQAE 876
 Db 751 PGLGQC-----EVLTEQESMSLQOMDSLIGDLHQEAQINQK----- 788
 QY 877 EERIRKHAARAEKAPQEGIAKAKLVMRTPITGTDENHRYMLF--SDEVPGLFIEKGMWH 934
 Db 789 -----IHTGLKIRFPPTGTRFHRNYWMLAHTDKV-----IIESLAT 826
 QY 935 DSIDYRFNHHCQHTVSGDEYDCPRSKKANLGKASNMNTQHGTAETEVAVETT----- 986
 Db 827 TSVN--NPACNANEYAKDP--PTLEQRVPGACETIDLDVIACVEDLVDDVLLRAKAD 881
 QY 987 ---TPKQGNL-----WFLCDSQKELDELNCLHPQIGRESQLKERLEKRYQDIHSHI 1037
 Db 882 KKTRKRYRRIENHMKRGWMTQNRDCVESLASCMLSRGIRERALLRLTTKPW----- 933
 QY 1038 LARPNLGLKSCDGNQELNLFRLSLIEVATRLQKGLGVYEETSEFEARVLSLEKLKDF 1097
 Db 934 -----FLNELKGTITIEPVGEKSDLELVRQGWTRINT-AIDKL----- 972
 QY 1098 GECVIALQASVIKKPLOGF-----MAPKQRRKLQSEDSA-----KTEEVDE 1139
 Db 973 -QC--HLAKMSDVSKPLPSITPTEQKPIVVPPTWALAQIVKDDMAWKVIDEEDVQDELDE 1029
 QY 1140 ---EKQVTEAKV-----ASALEKWTAI--RGAQTFSRMHVLLGLMDACIKWDMSENARC 1191
 Db 1030 TIIRQKIITADMDQDTCQLFEDWKSYVSTAQTTSQLMVALQTLLEGIMMWSREALC 1089
 QY 1192 KVCPPKGGEDDKLILDCENKAFHLFCLRALPYEVPDGEWQCPACOPATARRN----- 1243
 Db 1090 QIC-KSMGDGEMLVCDGCEGCHMECFRPMTKVPEGWFCQRCREEKSGRPMCMFCSRE 1148
 QY 1244 -----SRGRNYTBEASDESDEDEE-----EEEEEE 1274
 Db 1149 TGNLHQCRCAVHVHQESQDGPKEAIPETFIGHCQEMQMRFVKRLILRSESEREL 1208
 QY 1275 BEEDYEVAGLRPRKTTIRGKHSVIPPAARG---RRPGKPHSTRSQPKAPVDDADV 1331
 Db 1209 EDDNHAENG-----ENTKNGHMGNGMGAIAIGVHNQNGVKGNLKRK-----LEV 1253
 QY 1332 DELVLOTKRSRRQSLELQKCEIILHKIVKYRFSWPFREPVRTRDEADYVDVITHMPDQ 1391
 Db 1254 PSI-----GGLPKNKELCOLMDELUVQANALPFLBPVNPVLVPGVKMIISKPMDLK 1307
 QY 1392 TVQNKSCGYSYRQVEFLTDMKQVFTNAEVVNCRCGSHVLSWVKTEQCLVLLHLKLP--- 1449
 Db 1308 TIROKNEKLIYETPEDFAEDIELMFANCRQFNIDHSEIGRAGIS-----LHKFFQKR 1359
 QY 1450 ---GHPVVRKR-----KKFP-----DRL-----AEDEGDS-----PEAVGOS 1480
 Db 1360 WKQLKYNFTKRLRLHPKQYPIILHRHTCSETLSERLVAFAEVAGFAEQARTLGPAAAGSE 1419
 QY 1481 RDERRSREA 1490

Db 1420 ELEAEPSSTA 1429
RESULT 5
T12495
hypothetical protein DKF2p3434H071.1 - human (fragment)
C:Species: Homo sapiens (man)
C>Date: 23-Jul-1999 #sequence_revision 23-Jul-1999 #text_change 09-Jul-2004
C:Accession: T12495
R:Poustka, A.; Klein, M.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, June 1999
A:Reference number: Z17525
A:Accession: T12495
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-449 <POU>
A:Cross-references: UNIPROT:Q9UIF8; EMBL:AL080173
A:Experimental source: adult testis; clone DKF2p3434H071
C:Genetics:
A:Note: DKF2p3434H071.1
C:Superfamily: transcription factor GCN5; bromodomain homology
F:366-421/Domain: bromodomain homology <BRO>
Query Match 4.9%; Score 388; DB 2; Length 449;
Best Local Similarity 23.3%; Pred. No. 1.7e-10;
Matches 111; Conservative 81; Mismatches 169; Indels 116; Gaps 12;
QY 1002 KELDALLNCHPGIRESOLKERLEKRYODIHSIHLAKPNGLKSCDQNELNPLFS 1061
Db 6 EDLKALKVLHURGIREKALQKIQK-HLDYITQACLKNK-DVAITELNENES--NOVTR 61
QY 1062 DLIE--VATRLQKGLGYVEETSEFEARVISLEKLKDFECVIALQASVIKFLQGFMA 1118
Db 62 DIVENWSVEEQAMEMDLVLOQVEDLERVAS-----ASLQ-----VKGWMC 103
QY 1119 PKQ-----KRLQSEDSAKTEVEDEEK 1141
Db 104 PEPASEREDLVPEHKSFTKLCKEHDGEFTGEDSAHALERKSDNPLDIAVTRLADLER 163
QY 1142 KNVE--EAKVASALEKWKTAIREAQTFSRMHVLGLMDACIKWDMSEANARCKVCPKGE 1199
Db 164 NIERLEEIDAPGLRVRRALSEARSAAQVALLCQLOKSIWKSIMKYCOICRKGN 223
QY 1200 DDKLILCDENKAFHLFCLRPALYVEPDGEWOCAPACATA-----RRNSGRNYTEE 1252
Db 224 BELLLLCDCGCKGCHYCHRPKITTPDGDWFCPACIAKASGOTLKIKLHVKGKKTNES 283
QY 1253 SASE-----DSEDDSEDEEEEEEEEEEEYEVAGLRPRKRTIRGKSHVIPPAAKSG 1306
Db 284 KKGKVTLAGDTEDEDS-----ASTSSS 306
QY 1307 RRPKKPHSTRSQKAPPVDDAEVDLVLQTKSSRSROSLELOKCEIILHKIVKRYFSW 1366
Db 307 LKRGNKDLKRRKMEENTS--INLSKQESFTSVKPKRDDSKOLALCSMLTETMETHEADW 364
QY 1367 PREPRTDEADYDYVITHPMDFTQVQNKCSGYSRVSQVEFLTDMKQVFTNAEVN 1423
Db 365 FELLPNVLKVPQKVKVKKPMDFSTIREKLSGGQYPNLETTALDVLRLVFDNCETFN 421
RESULT 6
S64146
probable membrane protein YGL133w - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein G2842
C:Species: Saccharomyces cerevisiae
C>Date: 17-May-1996 #sequence_revision 17-May-1996 #text_change 09-Jul-2004
C:Accession: S64146; S71739
R:Escibano, V.; Eraso, P.; Portillo, F.; Mazon, M.J.
submitted to the Protein Sequence Database, May 1996
A:Reference number: S64144
A:Accession: S64146
A:Molecule type: DNA

A:Residues: 1-1264 <ESC>
A:Cross-references: UNIPROT:P53125; EMBL:Z72655; NID:g1322701; PID:e243361; PID:g132270
A:Experimental source: strain S288C
R:Escibano, V.; Eraso, P.; Portillo, F.; Mazon, M.J.
Yeast 12. 887-892, 1996
A:Title: Sequence analysis of a 14.6 kb DNA fragment of Saccharomyces cerevisiae chromo
A:Reference number: S71733; MUID:96437978; PMID:8840506
A:Accession: S71739
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1264 <ESW>
A:Cross-references: EMBL:X92670
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1995
C:Genetics:
A:Gene: SGD:ITC1
A:Cross-references: SGD:S0003101
A:Map position: 7L
C:Keywords: transmembrane protein
F:549-565/Domain: transmembrane #status predicted <TM>
Query Match 4.3%; Score 345.5; DB 2; Length 1264;
Best Local Similarity 19.5%; Pred. No. 4.6e-08;
Matches 271; Conservative 206; Mismatches 492; Indels 423; Gaps 58;
QY 3 PLLGRPFPLVNPPLPGEEPPFTIPHTQEAFFRTRREEYEARLERYSERIWTCKSTGSSQLTH 62
Db 8 PILLPDPKL--PLDLNVQVWHIEETGEWFSVVEEFLERDFYTRHHFTCEITGTSLTTF 65
QY 63 KEAWEEOEVAELLKEEPAPYKLVLEMHUN-TASLEKLVDTAWLEIMTKYAVGEED 121
Db 66 FOALDSEETQFYKVEEFLPLKREPVARFLHFGIRRLDALVEKVARFNKDFPFG-- 122
QY 122 FEVGKEMLVKIVKHPLEKVDDEATEKKSGDACDPSDDKENS--QIAQDHQKE 177
Db 123 --VYLKQKDSSTSSNQOSTPQDDMVEINSVGNPLGQYQVQR 168
QY 178 TVVKEDEGRRESINDRARRSPKLPSTSLKGERKWAPPFLPHKYVDKVLQNEDKIISN-- 235
Db 169 YVKEKQVFNATINPESREIVPAHT-----KY---MLIEEAASNKS 208
QY 236 --VPADSLIRTRPNKEIVYFIRHNALRAGTENAPWVVEDELVKYKSLPSKFSDFLL 293
Db 209 FIVDQCIYRDRSTETKHLIKCFKITTQIRASSKMGAPWCVKPEYLA MYGLTMEWPK-- 265
QY 294 DPYKMTLNPSTKTK-----NTGSPD-----RKPSKSKTSSSLSSPLNPKLWCHVHL 342
Db 266 DMLKYEDEFPVARRSNSANVSPSEKKNKRSKSGKSNSTNDASNKKEKTK-- 318
QY 343 KKSLSGSLPKVKNSKNKSPKEEHLSEMMKMSPNKLHTNPHIPKGGPPAKPKGHSKDKPL 402
Db 319 ----KRPTEVNDSENNSEED-----KKGQNTSET 347
QY 403 KAKGRSGKILNGQKSTGNSKSPKGLKTPKTKXKQMTLL-----DMAKGTQKMTAPRNS 457
Db 348 HSKGRKKE--ANEENPTENVES---VPTANAEPQAVTITSIMDILA----- 389
QY 458 GGTPTSSKPHKHLPPAALHLIAYYKENKDRKRSALSVCVSKTARLLSSSDRBLPEE 517
Db 390 -----LPYQH--PPNIPFNITYNEK-----LECI-----SLGSTKLSRPFDS 425
QY 518 LRSVLQKRYELLE---HKKRWASMSBEEQRKEYLK-----KKREEL 554
Db 426 FGKLLQ-AQOFLNTFGSKICLHFSLDQFITSKCTDPYELKGEVVLNIRVQTSTKEQEI 484
QY 555 KKK---LKEKAKERREKEMLERLEKQRYEDELGTGK---NLPAFLV-DTPBG---LPN 604
Db 485 ENNGLPMKNKATTEEDSENPSDWQRNSFIRDMIMKRNSDKVEYKIVHDDPASDDILDN 544
QY 605 TLFQDVAMVVEFLSCVSGLLLPDAQYPITAVSLMEALSADKGGFLYLNRLVILLQTLLO 664
Db 545 INHNGSALLIEVFTALLRLFI-NEEGDWSCIVVENWIIDDKG-----VLMER--- 590

```
QY 665 TLLQDEIADYDYGELGKMLSEIPLTLHVSSELVRL--CLRR--SDVOERSEG-SDTDDNKO 719
Db 591 ---KURGE--GEAQKRNAGHYFLQDKEKINLQKTLKENATEVQKESDAKNETNSGD 645
QY 720 SAAFDNEVDDEFLEK-----LETSEFFELTSE-- 747
Db 646 SKSDSDSEERDPKLEKCLNRYNNVNIERTKQFNNSYWLIIILGLVLEDCRHLPHYTEFI 705
QY 748 ----EKL-----QILTALCHRI-----LMTY-----SVQDHM 770
Db 706 DSFIEKIIPKOISATQLPKQLWRNFCRLSFDKVNALWILVDLVSHFSPDIIKAAVDDSM 765
QY 771 ETRQOMSALWELAVLKEENDKKAERQKEMEA-----KUKENGKVENGLGKTD 824
Db 766 E---LCGQIRSERFKVARELKTEAAVLSNLQGLQAOEKLKNTENTPISADGADKKD- 820
QY 825 KKRIKVPQVQTEAEDMISAVKSRRLLAIAQAKEREIQEREMKVKLERQAEERIRKH 884
Db 821 -----DSESENSEPIDLIIIEKKQKLEEQDKVQALQ----- 852
QY 885 AAAEKAFQEGIAKAKLVMRTPIGTDNRNRY-WLFSDEVP-----GLFIEKGWHD 937
Db 853 --SDKNFLDNCLEFQDLQRLKPLGLDRYCNRYFWDHNGVPPQYPAGM-NETPKSNNSL 909
QY 938 DY-----RNNHCKDH-----TVSGDDEY 956
Db 910 SVHSRLLIQGPKASAKPFLNVSDEQLSNWQKINSEGISATREVFISKTSSGSYNY 969
QY 957 CPRSKKAN-LGNASMNTO-HGTATEVAVETTPKQ---GQNLWFLCDSQKDELNLNCL 1011
Db 970 VENGIEVELLSDNRVNPILIELTPQKKIMDETSPSRLLSPDWICYDKLEDLRIMDWL 1029
QY 1012 HPQIRRESQLEKREKRYODIHSIHLARKPNLGLKSCDGNQELL-----NFI 1060
Db 1030 DNWGRKEHDLRQIRPIMERIKSSLSL-RDHALSHTAFTKNEBKLLKELENNEFTENELN 1088
QY 1061 SDLIEVATRLQGLGYVEETSEFPAVLSLEKLDFGECVIALQASVIKFLQGFMAPK 1120
Db 1089 VDSMDVDDK--NSGV-----KSEVDVQVDAEBK-----REAVIDEKLEVIAD 1130
QY 1121 QRRRLQSEDSAKTEBVE-----DEEKWVEE-----AKVASALEKWKTAIR 1161
Db 1131 ----MKLDDSSKTRNVLNRIOELEDQORDELLEQKKSINSQRPGARILARSRKTKIS 1185
QY 1162 EAQTSRMRHVL 1173
Db 1186 RGNKVNKQIEIL 1197

RESULT 7
T34418
Hypothetical protein F12F3.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T34418
R:Puliton, B.; Wohldmann, P.
submitted to the EMBL Data Library, July 1998
A:Description: The sequence of C. elegans cosmid F12F3.
A:Reference number: Z21521
C:Genetics:
A:Status: Preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-3488 <FUL>
A:Cross-References: EMBL:U80022; PIDN:AAC35885.1; GSPDB:GN00023; CESP:F12F3.3
A:Experimental source: strain Bristol N2; clone F12F3
C:Genetics:
A:Gene: CESP.F12F3.3
A:Map position: 5
A:Introns: 281/3; 332/1; 562/3; 600/3; 1866/3; 1944/3; 3393/1

Query Match 4.1%; Score 325.5; DB 2; Length 3488;
Best Local Similarity 19.1%; Pred. No. 1.2e-06;
Matches 334; Conservative 230; Mismatches 619; Indels 567; Gaps 75;
```

```
QY 36 EYEYERLERYSERIW-----TCKSTGSQLTHKEAWEEEQEVAELLKEFPAPW-YEKLVL 89
Db 280 DRMEYRHEDEWRKXWLIILKIDCKO-----EEAEYACQAINVAGEAWCSDVVV 327
QY 90 EMVHH---NTASLEKLVDTAWLEIMTKYAVGEECDFEVGKEKMLVKV-----IVKIHP 139
Db 328 HMBESRRDDKSVDEVDSTVLE--EKDDGDDKSKPKTKKIIKKKETPSEQVTAAP 385
QY 140 ----LEKVEDEATEKSGACDPSKSSQIAODHOKKETVVVKEDEGRRESINDRAR 195
Db 386 EOOKISEVDVQSVABTEVGAKKCPDAEKPTDLSAKAKDSKSK--KSDPEASTSEKSTT 442
QY 196 RSPKRLPTSLKKGKRWAPPK--FLPHKYDVKLQNEDKIISNVPADSLIRTERPP----- 248
Db 443 EKPNTDKTSKSAEKKTVPKKEVTGKPLEAKKPVEDKXDAQSPSS--KSSPPTDGKK 500
QY 249 NKEIVRYFIRHNALRAGTGENAPWVVEDELVKYKSLPSKFSDF----- 291
Db 501 KQIIPKALF-----IPDEISRFGDPSTMHSETNITTTIRREGSADAK 544
QY 292 --LLDP-----YKMTLNPSTK-----RKNTGSPDRK----- 316
Db 545 TPLVEPLSASVMKVFYTLVESAKAEPSFKRRSETPDDKSRKKEGLPPAKKSEKDEV 604
QY 317 -----PSKSKTDNSSL--SPLNPKLWCHVHLKKSLSGSPKVKNSKNSKSP 364
Db 605 AEKOSTEALIESKKEVDESKISBQQPSDKN-----KSEVVGPEKAAAGPETKKD 657
QY 365 HLEEMKMSPNKLHNFHIPKGGPPAKKPGKHSKDKPLKAKG----- 406
Db 658 -IEE-----VPKKTIIKKTEK--SDSSISQKSNVLKPADDOKSKSDVD 700
QY 407 RSKGLNQKSTGNSKSPKGLKTPKTQKQWTLTD-----MAKGTQKMTAPRNSGG 459
Db 701 KSKKTEDQTKVATDSKLEKAADTK-QIETETVDDSKKKVLKKTSEKDSFISQKSE 759
QY 460 TPRSSSKPHKLPAPALHLIAYYKENDRED-----KRSA-----LSCVIS 500
Db 760 TTPV-VEPTKPAESEAQIAEVNKAQKQEVDDNLKREAVAAKIADEKLIKIEANIK 818
QY 501 KTAALLSSEDRARLPEELR---SLVQKR--YELLEHKRWASMSBEQKRYLKKRE-EL 554
Db 819 KTAVEAAKQKQKDEQKLETEVVSKSAAEKLEKQ-AQIKKAAEADAVKQKELNE 877
QY 555 KKKLKEKAKERREKEMLER-----LEKOKRYEDQELTGKNLPAPFLVDTPGLPNT 605
Db 878 KNKLEAAKSAADKCLKLEESAASKKVSESVKPGEEKTKAGKKTQVVESEPTSKTI 937
QY 606 LFGDVAMVVFELSCYSGLLLPDAQYPITAVSLMEALSADKGGFLYNRVILLQTLQ 665
Db 938 DTKOV-----GATEPADETPKKIITKKTSEKSDSS----- 967
QY 666 LLQDEIADYDYGELGKMLSEIPLTLHVSSELVRLCLRRSDVQSESGSDTDNKO 725
Db 968 -ISQKSATDSKVSQKQEQDEPTKPAVSE-----TQMTADAKSKKQETDEK 1018
QY 726 NEVQDEFLEKLETSEFFELTSEKLIQILTALCHRLMTYSVQDHMETQOQSAELWKE-R 784
Db 1019 AEIAAK--TKQEADEKSKLDAQEKIK-----KVSEDDAARKEKELNDKLESE 1065
QY 785 LAYLKEENDKKAERQKQ---KREKEMAKNKENGKVENGLGKTDKRRKRVKPEPQV 840
Db 1066 IATKASADKCLKLEBQAKKAAAEVAAAKQKQKDE-----QLKIDTAA 1110
QY 841 DMISAVKSRRL-----AIQAKKEREIQEREM-----KVKLERQAEERI 880
Db 1111 SKKAAAEKLEKQKQAIKKAAGADAVKQKLEKKNLEANKKSAAKLIKIE---EESA 1167
QY 881 RKHKA AEKAFQEGIAKAKLVMRTPIGTDNRNRYWLFSEVPGFLFTEKGNVHDSID 940
Db 1168 KSKQTVBEQAKLDAQTKAKTAEKQTKLEK-----EKS----- 1200
```

QY 941 FNHCKDHTVSGDEDVCPRSK--KANLGRKASNMNTHGTATEVAVTTTPKQG--QNLWF 996
 Db 1201 ----TKESKSTVDEKPKKKVLKKTETKSDSISQKSETSKTVESAGPSESTQKVAD 1256
 QY 997 LCDSQKELDELLNCLHPQIGRESQLEKRYQDIIHSHILARKNGLGKSCDGNQELL 1056
 Db 1257 AARKQKETDE-----KQKEAEI-----TAKSADKSKUL- 1286
 QY 1057 NFLRSLDIEVATRLQKGLGVVEETSEFEARVISLEK-----LKDPGECVIAL 1104
 Db 1287 -----EAESKLKK-----AAEVEAAKQKEKEDQLKLDTEAASKAAAEKLELE 1330
 QY 1105 QASVIKFIQGFVAPKPK--RRKLQSEDSAKTEVEDEKQMVBEAKVASA-----LEKW 1156
 Db 1331 KQSHIRKAAAEVDAVKQKLEEKQRLSE--AATKKADEKLEKQKKAABIALIEIQ 1388
 QY 1157 KTAIRAQOTFSRMHVLGLMDACIKWMSAENARCKVCPKKGDD---KLILDCENKAP 1213
 Db 1389 KEQEKLAQOSRLDEAKSAEKQLESETSKQTEAPKESVDEKPKKVLKKTETKS- 1447
 QY 1214 HFLCLRPALYVDPGEWQCPACOPATARNRNGRNYTERESASDSDSDDESEEEEEEE 1273
 Db 1448 -----DSSISQKSKSAKSTVDAETLESDFNLVEKTVQKVE 1484
 QY 1274 EEEDEYEVAGLRP-----RKTIRGHSVIPPAAARSGRRPKKPHSTR 1317
 Db 1485 QSPDETSATIKRDPKPAKTEIEISKQDDGDKTTTQK---PPKPEDSEATPKKRVVK 1540
 QY 1318 RSQKAPPVDDA-----EVDE-----LVLTQKSS-----RRQSL 1349
 Db 1541 KTKQSDSVASDASLADVSKLSDVEKPKKVLKKTETKSDSISVSETSVDTIKPESVBI 1600
 QY 1350 --QKCEE--ILHKIVKYRFPPEPVTRDEAEYDVIITHPMDQTVQNKCSGVSRSVQ 1406
 Db 1601 PTEKAEQMLHN-----RFS-----TDSAVESEPNNAHKDDTE-----KTTD 1637
 QY 1407 EFLTDMKQ--VFTNAEYVNCGRSHVLSCMVKTEQCCLVLLHKLPHGVYVRKRKFPDR 1464
 Db 1638 DMWTRKSSAIFSDD-----QSISKTSSE-----GRR--RRRTGFASK 1676
 QY 1465 LAED-----EGDSEPEAVGSRDEDRSREABIQ--EWLQTSLSYSAKINSKQ----- 1510
 Db 1677 FASDTLALRGDN-----VEIEAEALLAEDDTVTWKVNGKQADLNSRCH 1719
 QY 1511 --HNCFMFVLV 1518
 Db 1720 MSHFTFRTLII 1729

RESULT 8
 S67593
 transport protein USO1 - yeast (Saccharomyces cerevisiae)
 N:Alternate names: protein D2552; protein YDL058w
 C:Species: Saccharomyces cerevisiae
 C>Date: 12-Jul-1996 #sequence revision 12-Jul-1996 #text_change 09-Jul-2004
 C:Accession: S67593; A38455; S30782
 R:Bloeker, H.; Brandt, P.
 submitted to the Protein Sequence Database, July 1996
 A:Reference number: S67587
 A:Accession: S67593
 A:Molecule type: DNA
 A:Residues: 1-1790 <BLO>
 A:Cross-references: UNIPROT:O07380; EMBL:Z74106; NID:g1431058; PID:e253003; PID:g1431059
 A:Experimental source: strain S288C
 R:Nakajima, H.; Hirata, A.; Ogawa, Y.; Yonehara, T.; Yoda, K.; Yamasaki, M.
 J. Cell Biol. 113, 245-260, 1991
 A:Title: A cytoskeleton-related gene, USO1, is required for intracellular protein transp
 A:Reference number: A38455; MUID:91185402; PMID:2010462
 A:Accession: A38455
 A:Molecule type: DNA
 A:Residues: 1-389, 'TA', 392-724, 'S', 726-1790 <NAK>
 A:Cross-references: GB:X54378; NID:g4777; PIDN:CRA38253.1; PID:g4778
 A>Note: the authors translated the codon ACT for residue 768 as Ile

R:Hostetter, M.K.; Herman, D.J.; Bendel, C.M.; McClellan, M.; Tao, N.; Kendrick, K.E.
 submitted to the EMBL Data Library, February 1993
 A:Description: An integrin analogue in Saccharomyces cerevisiae.
 A:Reference number: S30782
 A:Accession: S30782
 A:Molecule type: DNA
 A:Residues: 71-846, 'E', 848-923, 'K', 925-1252, 'I', 1254-1318, 'V', 1320-1460, 'S', 1462-1580,
 A:Cross-references: EMBL:L03188
 C:Genetics:

A:Gene: SGD:USO1; INT1
 A:Cross-references: SGD:S0002216; MIPS:YDL058w
 A:Map position: 4L
 C:Keywords: coiled coil; transmembrane protein
 F:326-342/Domain: transmembrane #status predicted <TM1>
 F:394-410/Domain: transmembrane #status predicted <TM2>
 F:617-633/Domain: transmembrane #status predicted <TM3>

Query Match 4.0%; Score 322; DB 2; Length 1790;
 Best Local Similarity 20.2%; Pred. No. 7.9e-07;
 Matches 274; Conservative 222; Mismatches 471; Indels 392; Gaps 58;

QY 17 PCGEPPFTTPTQEAFTREEYEARELRYSERIWTCKSTGSSQLTHKEAW-----EE 69
 Db 719 PDEEPINKI-----SFEVEKLQROCTKLKGBITSLQT--ETESTHENLEKLIALTNEH 771
 QY 70 QEVAE-----LLKEEFPANWEKLVLEVMVHNTA-SLEKLV-----DT 105
 Db 772 KELDEKYQILNSSHSLKENFS-----ILETELKNVRSLSDEMTQLRVLETYKXENQT 825
 QY 106 AWLEIMTKYAVGEEC--DPEVGKERMLVKVIKIHLEKV-----DEEATEKKSDG 154
 Db 826 ALLEYKSTTHKQDSIKTLEKLETTLSQKKAEDGINKWGDLPALSRMQAVEEN--- 882
 QY 155 ACDSPPSSDKENSQIAQHOKKETVVEDEGRRESINDRARRSPKLPSTSLKKGKRWAP 214
 Db 883 -CKNLQKEKXSN--VNHQKETSLKEDIAAKIT----- 913
 QY 215 PKFLPHKYDVKLQNEDKLIISNPADSLIRTERPPNKEIVRYFIRHNALRAGTGENAPVWV 274
 Db 914 -----EIKAINENLEEMKIQCNLSKEKEHISKELVEYKSRFOS----- 952
 QY 275 EDELVKYKSLPSKFSDFLLDPYKYMTLPSTKRNKTGSPDRKPKSKKTDNSSLSPPLNP 334
 Db 953 HDNLVAK--LTEKLKS--LANNYKDMQAE NESLIKAV-----ESKNESIQLSNLQN 1001
 QY 335 KLWCHVHLKKSLSGSLPKVYKNSKNK-----SPEEHLSEEMKQMSPNKLTNPHIPKGGP 390
 Db 1002 KI-----DSMS-----QEXENFQIERGSIEKNIQKKTISDLE-QTKBEIISKSDS 1047
 QY 391 AKKPGKHSKPLKAKGRSKGILNGQKSTGNSKSPKGLKTPKTKMQMTLLDMAGTKQM 450
 Db 1048 SKDEYESQISLKEKET-----ATTANDENVNKISLTKTREELEAEIAYKNL----- 1097
 QY 451 TRAPRNSGTP-RTSSKPHKLPPAALHLIAYYKKNKREDKRSALSVCISKARLSS- 508
 Db 1098 ----KNELETKLETSEKALKEVKEEHL-----KEEKIQLEKEA-----TETKQQLNSL 1143
 QY 509 ----EDRARLPPELSLVOKRYELLEHHKKRWAS-----MSEQRKEYLKKREEL 554
 Db 1144 RANLESLEKEHEDLAAQLKKYBEQIANKEROYNEEISQLNDEBITSTQGENESIKKKNDEL 1203
 QY 555 KKKLEKAKERREKEMLEKQKRYEDELGTGNLPAFLPVDTPGELPNTLPDGVAMV 614
 Db 1204 -----EGEVKAMKSTSEESQNLKXSEIDALMLQIKELKKQKQETNEASLESIKS-V 1253
 QY 615 EFLSCYSGILLPDAQYPIITAVSLME---ALSADKGGFLYLNRLVILLQTLQLQDSEI 671
 Db 1254 ESETVKIKELQDECNFKKEVSELEDKUKASEDKNS-KYLE-----LQKES 1298
 QY 672 AEDYELGMKULSEIPLTLHSVSELVRLCLRRSDVQSESGSDTDNKKOSAAAFEDNEVQDE 731
 Db 1299 EKIKELOAKTTTELKIQLEKITNL-----SKAKEKSESELRLKKTSS--EERKNASE 1349


```
QY 973 TOHGTA TEVAVETTTTPKOGONLWFLC-----DSQKELDELINCLHPGIGRESOLK 1022
Db 2232 SKAQTEVLOQKVCOTDLOGENKELLSQLEETHLYHSSONELAKL-----ESELK 2281
QY 1023 ERLEKRYQDI IHSIHLARKNPLGLKSCDGNQEL-----LNFLRSDL-----IEVATRLQ 1071
Db 2282 S-LKQOLTDLNSLECKEKGKGNLSEGIIRQOEADITNSKFSVEQLETDLOASRELTSLRH 2340
QY 1072 KGLGVVETSFBEARVISLEKLDGFCVIALQASVIKKFLOGFMAPKQKRKLQSBDS 1131
Db 2341 -----EEINKEQKIIISLGSKE-----EAIQVIAELRQOHKEIKELLENLSQSEEE 2388
QY 1132 AKTEEVDEKQVVEAKVASALEKWKTAIRE-----AQTFSRMHVLLGLMDACIKWMSA 1186
Db 2389 ENIVLEENKKAVD--KTNQLMETLTKIKKENIQKAAQ-----LDSFVKMSGSL 2435
QY 1187 ENARCKVCPPKGGEDDKLILCDENKAFHLFCLRPALYEVDPDGWQCPACQAPATARRNSRG 1246
Db 2436 QNDR-----DRIV-----GDYQ-----2447
QY 1247 RNYTESASEDESDEEBEEDDEEBEEDDEEBEEDDEEBEEDDEEBEEDDEEBEEDDEEBEED 1306
Db 2448 -QLEERHLISILEKQOLIQEAAAENKLEKE-----IRGLRSHMDL-----2488
QY 1307 RPPGKKPHSTRSQPKAPPVDDAEVDLVLQTKRSRRQSLE--LQKCEEIILHKIVKYRF 1364
Db 2489 -----NSENAKLDAELIQYREDLNQVITIKQOQKLEVLQOQKNELENKAKYLEE 2540
QY 1365 SWPFREPVRDEAEDYDVI THPMDPQTQVQNKCSGVSRSVOEFLTMKQVFTNAEVNVC 1424
Db 2541 KLKESSEANEDLRFSNALQESKQDLS-----KEIESLKVISISQLTRQVTAQOE 2589
QY 1425 RGS-----HVLSCWKTEQCLVVLHKKHL-----PGHPYVRRKXK 1459
Db 2590 EGTGLGYHAQLKVEEVEHRLSALFSSQKRIAELEEELVCVQKEAAKVGES--IEDKLK 2647
QY 1460 KPFDRLAEDG-----DSEPAVQO-SRD-----EDRRSREAEIOEWLQDTS 1500
Db 2648 KELKHLHDAGIMRNTEETAEBERVAELARDLVEMEQLMLMVTKENKGLTAQIQSFGRSMS 2707
QY 1501 LYSAKINSKDH 1511
Db 2708 SLQ-----NSRDH 2715

RESULT 11
giantin - human
N:Alternate names: macrogolgin
C:Species: Homo sapiens (man)
C:Date: 19-Oct-1995 #sequence_revision 26-Jan-1996 #text_change 09-Jul-2004
C:Accession: A56539; S37536
R:Seelig, H.P.; Schranz, P.; Schroeter, H.; Wiemann, C.; Griffiths, G.; Renz, M.
Mol. Cell. Biol. 14, 2564-2576, 1994
A:Title: Molecular genetic analyses of a 376-kilodalton Golgi complex membrane protein
A:Reference number: A56539; MUID:94187728; PMID:7511208
A:Accession: A56539
A:Molecule type: mRNA
A:Residues: 1-3259 <SEE>
A:Cross-references: UNIPROT:Q14789; EMBL:X75304; NID:g405714; PIDN:CAA53052.1; PID:g4057
C:Genetics:
A:Gene: GDB:COLGB1; GCP; GCP371
A:Cross-references: GDB:454958
A:Map position: 3q13.31-3q13.31
C:Superfamily: giantin
C:Keywords: coiled coil; Golgi apparatus; transmembrane protein
P:3238-3254/Domain: transmembrane #status predicted <TMN>

Query Match 3.8%; Score 302.5; DB 1; Length 3259;
Best Local Similarity 18.5%; Pred. NO. 1.2e-05;
Matches 358; Conservative 279; Mismatches 633; Indels 661; Gaps 80;
```

```
QY 30 EAFRTREYEARLERYSERIWTCKSTG-----SSOLTHKEAWEEQEVAELLKEEFPFA-- 82
Db 1031 ETERGEVEEDKENKEYSEKCTSKQOEIYVLKQTISEKV--ELQHRKQLEEKLAAREE 1088
QY 83 WYKLVLELM-----VHNTASLEKLV-----DTAMLIEIMTKYAV 116
Db 1089 QFOALVKMNQTLQDKTNQIDLLQAEISENQAIQKLITSNNTDASDGDGVAL-VKETVVI 1147
QY 117 GEECDFEVG-----KEKMLKVIKVIHPLKVDDEATEKKS--DGACDSPSDKEN 165
Db 1148 SPPC---TGSSEHWKPELEEKILALEKEQOLQKQLQALTSRKAILKXAKQEKERHLEE 1204
QY 166 SSOIAQDHOKKETVVKEDEGRRESINDRARRSPKLTSLKKGERKWKAPPKFLPHKYDVK 225
Db 1205 LKQKDDTNRLOEQPDEQSKENENIGDOLRQLOQVRESI-----DGK 1247
QY 226 LQNEED--KIISNVPA-----DSLIRTPPPNKIYVYFIRHNALRAG 265
Db 1248 LPSTDQOQSCSSTPGLEBPFLFKATEQHHTQPVLESNLCPDWPSHSEDA-----SALQGG 1301
QY 266 TG-ENAPVWVEDELVKYSLSPKSPDFLLDPKYKMTLNPSTKRNKTPGSPDRKPSKSKTD 324
Db 1302 TSVAQIKAQKLEIEAEKVELELKVSS-----TTSSELTCKSEVFOLOEQINQKQGLE 1352
QY 325 NSSLSPLNPKLWCHVH--LKKSLSGSPLKVYKNSKNSKSPPEHLEEMKMKMSPNK---- 377
Db 1353 IESLTKVSHS---AEVHAESIQKLESSQLQIAGLEHLRELQPKLDELQKLISKEEDVS 1409
QY 378 -----LHTNFHIPKKGPPPAKPGKHSK-----400
Db 1410 YLSGQLSBEKAALTKIQETIIEQEDLIKALHTQLEMQAK-----BHDERIKQLQVEL 1461
QY 401 -PLKAKGRSGILNGCKSTGNSK-----SPKGLKTPKTKMKQMTLLDMAKGT-QKWTR 452
Db 1462 CEMKQKPEEIGEEESAKQOIQRKQALISRKALKENKSLOEUSL---ARGTIERLTK 1518
QY 453 APRNSGGTPRTSGPKHKLPPAALHLIAVYKENKOR-----EDKRSALSQVSKT 502
Db 1519 SLADVESQVSAQNK-----KDTVLGRLLAQEERDKLITEMDRSILLENQSLSSCESLKL 1574
QY 503 ARLLSSEDRARLPEELRSL-----VQRYELL-----EHKKRWA-----536
Db 1575 ALEGLTDEKLVKEIESLKSSTAESEWQEKHKLQKEYEILLQSYENVSNSEARIQH 1634
QY 537 ---SMSEORKEYLKKKEBELKKLKYAKARREKEMLERLEKQRY-----580
Db 1635 VVEAVRQEKELYGLRSTEAANKKETEKOLOEAEQEMEEMKMKKFAKSKQOKTILELEE 1694
QY 581 EDOELTGKNLPAFLVDTPEGLPNTLFGD-----VAMVVEFLS-CYSGLL-----624
Db 1695 ENDRLRAEVHFA---GDTAKECMETLLSSNASMKEELERVKMEYETLSKQFSLMSKDS 1751
QY 625 ----LPDAQYPI-----632
Db 1752 LSEEVQDLKHQIEDNVSKQANLEATEKHNDQNTVTEEGTQSGTSGTSEEDSLSMSTRPTC 1811
QY 633 -----TAVSLMEALSADGGFLYLNRLVILLO-----TLLOTL-----666
Db 1812 SESVPSAKSANPAVSKDFSSHDEINNYLQOQIDQLKERIAGLEEBEKKNKFSQTSQENENK 1871
QY 667 -LQDIAEDYGBGLMKLSEIP-----LTLHSVSELVRLCLRRSDVQSESGSDTDNKSQA 721
Db 1872 TLLSQISTYDGLKMLQBEVTKMNLNQIQEELSRVTKLKTAESEKD--DLEERLMNQ 1929
QY 722 AFE-----DNEVQDFLEKLETSEFFEL-----TSEEKQILTALCHRIILM 762
Db 1930 LAELNGSIGNYCQDVTDAQIKNELLE---SEMKNLKKCVSELEBEKQOL-----1975
QY 763 TYSVQDHMETROQMSAELWKEKRLAVJ-----KEENDKKRAE-----KQKKEWEAKNKEN 812
Db 1976 -----VKEKTKVSEIRKYELEKIQAQAKPEPGNKSHAKELQELLKKEQKQVLOKQD 2028
QY 813 GKVENGLGKTRKKRIVKVF---EPQVDTE--AEDMISAVKSR-----849
```

Db 2029 IRYQEKISALERTVTKALEFVQTESQKDLTTKNTLAQAVEHRKKAQAEALASFVKVLLDDTQ 2088
QY 850 ----RLLAIOAKKEREIOBERMKV-----LER-----QAEERIRKHKAIAAKAFQ 893
Db 2089 SEARVLADNLKLELQSNKESVSKQKQEDLERLEQAEELKHEKQKQWQKL--D 2146
QY 894 GIAKAKLVNRRPTIG-----TDRNHNRYWLFSDVEVPGLFIEKGWVH 934
Db 2147 ALRREKVHLEET-IGBIQVTLNKKDKKEVQLOENLDSTVTQLAAFTKSMSSLDQDRDVI 2205
QY 935 DSI---DYEFNH--HCKOHTVSGDEDYCPRSK-----KANLGK-----NASNV 972
Db 2206 DEAKWKERFSDAIOKSEIEIRLKEDNCVSLQDLQRMQSIHMEELKINISRLHEHDKQIWE 2265
QY 973 TQGTGATEVAVETPTTPKQONLWFLC-----DSQKELDELINCLHPQIGRESOLK 1022
Db 2266 SKAQTEVQLQKVCOTLQGENKELLQLEBTRHLHYSSQNELAKL-----ESELK 2315
QY 1023 ERLEKRYQDIHSHIHLAKPNLGLSKDCGNQEL-----LNFRLSDL---IEVATRLQ 1071
Db 2316 S-LKQDLTDLNSLEKCKEKGKQNLGIIIRQGEADIQNSKFSVEQLETLQASRELTSLH 2374
QY 1072 KGLGLVVEETSEFARVISLEKLDKQFGEVIALQASVKKFLQGWAPKQKRLQSEDS 1131
Db 2375 -----EENKKEQKIISLSGKE-----EAIQVAIAELRQOQHDKEIKELNELLQSEBE 2422
QY 1132 AKTEEVDEKQKVEBAKVASALEKWKTAIRE-----AOTFSRMVLLGLMDACIKWMSA 1186
Db 2423 ENIVLEENKKAVD--KTNQLMETLTKIKENIQKQAO-----LDSFVKMSGL 2469
QY 1187 ENARCKVCPKGEDDKLLCLDCENKAFHLFCRLPALYEPDGEWQCPACQATARRNSRG 1246
Db 2470 QNDR-----DRIV-----GDYQ-----2481
QY 1247 RNYTESASEDSEDESEDESEDESEDESEDESEDESEDESEDESEDESEDESEDESEDESE 1306
Db 2482 -QLEERHLSIILEKQDLQTOEAANNKLEB-----IRGLRSHMDL-----2522
QY 1307 RRPKKPHSTRSQKAPPVDDAEVDELVLOTKRSSROSLE--LQKCEIILHKKVYRYF 1364
Db 2523 -----NSENAKDLAEIQTREDLNQVITIKDSQKQLLEVLQOONKELENKVKALEE 2574
QY 1365 SWPFPVTRDAEYDYVDITHPMDFTQVQNKCSGYSRSVQFELTDMQVFTNAEYVNC 1424
Db 2575 KLKSESEANEDLRRSFNALQEEKQDLS-----KEISLKVSIQLTRQVTLQ 2623
QY 1425 RGS-----HVLSCWKTEQCLVLLHKL-----PGHPYVRRRK 1459
Db 2624 EGTGLYHAQLKVKEEVEHRLSALFSSQKRIAELEELVQVQKAAKKVGE--IEDKLK 2681
QY 1460 KPFDRLAEDG-----DSEPEAVGQ--SRD-----EDRRSREABIQEWLQDTS 1500
Db 2682 KELKHLHDDAGIMRNTEETAEBERVAELARDLVEMEQLLMVTKENKGLTAQIQSFGSNMS 2741
QY 1501 LYSAKINSKDH 1511
Db 2742 SLQ-----NSRDH 2749

RESULT 12
smooth muscle myosin heavy chain 1 - mouse
C:Species: Mus musculus (house mouse)
C:Date: 10-Jun-1997 #sequence_revision 18-Jul-1997 #text_change 09-Jul-2004
C:Accession: JCS420
R:Haegawa, K.; Arakawa, B.; Oda, S.; Matsuda, Y.
Biochem. Biophys. Res. Commun. 232, 313-316, 1997
A:Title: Molecular cloning and expression of murine smooth muscle myosin heavy chains.
A:Reference number: JCS420; MUID:97242182; PMID:9125171
A:Accession: JCS420
A:Molecule type: mRNA
A:Residues: 1-1972 <Has>

A:Cross-references: UNIPROT:008638; DDBJ:D85923; NID:g1945077; PIDN:BAA19690.1; PID:g19
A:Experimental source: smooth muscle
C:Comment: This protein plays a role in smooth muscle cell contraction.
C:Superfamily: myosin heavy chain; myosin motor domain homology
C:Keywords: nucleotide binding; P-loop
F:88-771/Domain: myosin motor domain homology <MMOT>
F:178-185/Region: nucleotide-binding motif A (P-loop)

Query Match 3.7%; Score 298.5; DB 2; Length 1972;
Best Local Similarity 19.4%; Pred. No. 1e-05;
Matches 280; Conservative 233; Mismatches 533; Indels 401; Gaps 60;

QY 12 LVNPLPGEBEPFTTIPHQTQAFPT-----REYEARLERYSERIWTCKSTGSSQLTHKEA 65
Db 750 MIKALELPNLYRIGQSKIFFRTGVLAHLEERDLKI---TDVIMAFQCMRGYLARKAF 806
QY 66 WEEEOEVAELLKEEPANVEKLVLEVMHNTASLEKLVDTAWLEIWTK-----YAVGEEC 120
Db 807 TKRQOOLT-----AMKVIQRNCAAYLKLNRQWMLRFTKVKPLLVQTRQEE 852
QY 121 DFEVGKEMLVK-----KIVKIHPLEKVDDEATEKKS-----DGACDSPSDDKENSQ 168
Db 853 ENQAKSEEMOKITERQQAETELKEHQHTQLAEKTLLOEQLOAETSLYAESEEMVR 912
QY 169 IAQDHQKXETVVKDEGRRESINDARRSPRLPTSLKKGKRWAPPKFLPHKYDVKLQ 228
Db 913 LAAKQOELEILHEARLEERDRQ-----LQAEK-----KMAQQMLDLEEQ 959
QY 229 EDKIISNPADSLIRTERPPNKEIVYFIRHNALRAGTCENAPVVEDELVKYSLPSKF 288
Db 960 EE-----EEAARQKLQ-----LEKVTAEAKIKLEDDILVMDQNSKL 997
QY 289 S---DFLLDPYKYMTLNPSTKRNKGTSPDRKPSKSKTNDSSLSLSPNPKLWCHVHLKKS 345
Db 998 SKERKLLERVSDDLTTNLAEEKA-----KNLTKLKSKHESMISEL-----EVLKKE 1046
QY 346 LSG-----SPLKYVNSKNSKSPPEHLEEMMMOMSPNKLHTNFHPIPKGPPAKPGKHSXP 401
Db 1047 EKSROLEKLRKLEGDASDFHEQIADLQAQIAELKM-----QLAKKEEELQAALARLDEE 1102
QY 402 LKAKRS--KGI-----LNGQKSTGNSKSPK-----GLKTKTKWKQMTLID 442
Db 1103 TAQKNALKKIRELEGHISDLOEDLDSERAANKAKKQKRDIGEELEALKTELED--TLID 1160
QY 443 MAKGTQKVTAPRNSGGT-----PRTSSKPHKLPPAALHLIAYYKENKOREDKRSAL 495
Db 1161 -STATQOELRAKBEQEVTVLKALDEETSHSAQVQEMRQKHTQAVEELTEQLEQKRA- 1218
QY 496 SCVISKTARLLSSEDRARLPEELRSLVQKRYELLEPHKKRWASMSBEQRKEYLKKKKEELK 555
Db 1219 KANLQSKQTLKEN-ADLAGELRVLQAKQ--VEHKK-----KLEVLQDLQ 1265
QY 556 KKLKAKERREKEMLEKRYEDQELTGKNIPLAPFLVDTPEGLPNTLFGDVAWVVE 615
Db 1266 SKCSD--GERARAELSDKVHLQ--EVESVTG-----MLNEAEGKAIKAKDVASLGS 1315
QY 616 FLSCYSGLLLPDAQYPIITAVSLMEALSADKGGFLVNLRVILLQTLQLQDETAED- 674
Db 1316 QLQDTQELLQEBETROKLVNVTKLQJEDERN-----LQDQJDEEM 1356
QY 675 -----YBELGMKLSIPLTHSVSELVRLCLR-RSDVQSESESGSDTDKNKSAAFE 724
Db 1357 EAKQNLERHVSITLNIQLSDSKKLQDFASTIEMVEEGKRLQKEMEGLSQQVEEKAAYD 1416
QY 725 -----DNEVQDEF-----LEKLTSEFPFELTSEBKQLITLALCHRLMT 763
Db 1417 KLEKTNRLQOELDVLVDLQNRQLVSNLEK-KQKQFDQLAEEK-----NISSK 1466
QY 764 YSVQDHMTROQMSAEL-WKERLAV-----LKEENDKKRAEKKKEMAKNKENG 813
Db 1467 YA-----DERDKRAEAREKETKALUSARALEEALEKEELRTNMLKAEHEDLVSSKD 1521
QY 814 KVENGLGKTRKRIVKFEPQVD-----TEADMI SAVKSRRL-----LAIQAKKERE 861


```
Db 1628 ERRAEAGSKAADEEIRKQVMOEVDDELRAQDSEAAALNASKKTKSLVAEVDVEKEQL 1687
Qy 891 F0EGIAKAKLVMRRTIGTDRNHNRYWLFSDVPGFLFIEKGWVHDSIDYRFNHHCKDHTV 950
Db 1688 ED0ILAKDLVAKRALEVELEVRQLEEEE-----DSRSELEDSKRL 1732
Qy 951 SGDEDYCPKSKKANLGNKASMTQHTATAVETTPPKQGNLWFLCDQSKDELBNLC 1010
Db 1733 TTEVEDIKKKYDAEVQNTKLDKAKKLDV--DTLKQK-----LEDEKKKLN----- 1780
Qy 1011 LHPOGIRESQLERLEKRVQDIHSIHLARK-----PNLGLKSCDGNQELLNFLRSLIEV 1066
Db 1781 -----SERAKRLSENEEDFLAKDLAEVKNRSRAEKDRKYEKDLKDTKYKLNDEAAT 1833
Qy 1067 ATRLOQGLGYEETSEFARVLSLEKLDGFCVIALQASVTKKFLQGFMAPKQKRKL 1126
Db 1834 KQTEIGAALKLEDDIDELRSK-LEQOAK-----ATQADSKKTLG----- 1874
Qy 1127 QSEDSAKTEEVDEEKQWVE-EAKVASALEKWKTAIREAQTFRMHVLLGMLDACIKWMS 1185
Db 1875 -----EIDNLRQAIEDGKIKRLEKRA-----LE 1901
Qy 1186 AENARCKVCPKGEDDKLILDCENKAFHLFCLRPALYEVPGDGEWQCPAQATARNR 1245
Db 1902 GELEELRETVEAEADSK-----SEAEQSKRLVEL-----ELED-----ARRNLQ 1940
Qy 1246 GRNYTESASEDSEDESDEEBEEEEEEEDYEVAGLRL-RPRKTIIRGHKSHVPPAAR 1304
Db 1941 ----KEIDAKETAEDAKNLQREIVAEKGRLEESARTNSDRSKRL----- 1984
Qy 1305 SGRRPGKPKPHSTRSQPPVDDAEVDELVLQTKRSSRRQSLQKCEIHLKIVKYRF 1364
Db 1985 -----EABIDALTAQVDAEQAKAQKQIKENKKIETELKEYRK 2021
Qy 1365 SHPPFREPVRDEADYDVITHPMDPQTQVONKSCSGYSVQEFLLDMKQVFTNAEVNVC 1424
Db 2022 KF-----GESEK-----TKTKEFLVVE-KLETDYKRAKKEAADEQOQRLT----- 2060
Qy 1425 RGSVHLSVMKYTEQCLVLLKHLPCHPVYRRKRKFPDRLAEDEGSDPEAVGQSRDED 1484
Db 2061 -----VEND-----LRKHLSEISLL-----KDAIDKL-----QRDHD 2087
Qy 1485 RRSREABIQ 1493
Db 2088 KTKRELETE 2096

RESULT 14
JC5837
364K Golgi complex-associated protein - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 05-Mar-1998 #sequence_revision 13-Mar-1998 #text_change 09-Jul-2004
C:Accession: JC5837
R:Foki, C.; Fujiwara, T.; Sohda, M.; Hong, H.S.; Misumi, Y.; Ikehara, Y.
Cell Struct. Funct. 22, 565-577, 1997
A:Title: Identification and characterization of rat 364-kDa Golgi-associated protein re
A:Reference number: JC5837; MUID:98093490; PMID:9431462
A:Accession: JC5837
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-3187 <TOK>
A:Cross-references: UNIPROT:Q63714; DDBJ:D25543; NID:G516825; PID:BAA05026.1; PID:G5168
C:Comment: This protein plays a role in the formation and maintenance of the characterie
C:Superfamily: giantin
F:49-349,624-1176,1238-1707,1763-3114/Domain: coiled-coil leucine zipper #status predic
F:3165-3187/Domain: membrane anchor #status predicted <MAD>

Query Match 3.7%; Score 295.5; DB 2; Length 3187;
Best Local Similarity 18.3%; Pred. No. 2.4e-05;
Matches 351; Conservative 267; Mismatches 678; Indels 621; Gaps 67;
Qy 19 EBPFFTIPTQBAPRTREYEARLERYSRITWCKS--TGSSQLTHKEAWBEEQVAVELL 76
```

```
Db 919 EEPVCKEALQOELEMLRKESEQRKQLQAALISRKELLOQVSKLEELAKVREESTKDSL 978
Qy 77 KEFPAWYKLVLEVMHNTASLEKLVDTAW--LEIMTKYAVE-ECDPEVGEKMLVKV 133
Db 979 RES-----EKRELEDSKNKDDPEKYGTSEWELEVSLRLTITSEKEVELE-GIRROLKEK 1032
Qy 134 IV-----KIHPLEKVEEATEKSD-----GACDSSPSDKENS 166
Db 1033 AAAEELQALVQMTQDLQNKTKQIDLLQSEITENQATIKFITGTMDAGDGSVAKETS 1092
Qy 167 SQI-----AQDHQKQETVVKDEGR-----RESINDRARSPRK 200
Db 1093 VSSPPRAGGGEHWKPEL-----EGKIVDLEKEKTQKQLQALISRKAILKKAQEKKEH 1147
Qy 201 LPTSLLKKGKRWKAPPKFLPHKYDVKLVKLVKLVKLVKLVKLVKLVKLVKLVKLVKLVKLV 260
Db 1148 LKEELKEQDAY---RHLOEQFDGQSKENENI-----RAPLRQLOAKESTDQ 1191
Qy 261 ALRAGTGENAPWVVEDELVKYKSLSPSKFSDP-----LLDPYKY 298
Db 1192 QL-PGTGQOQEPHSGSEGLSEGTPEPASEDLHAAQPSHPGETATLQATVSVAQIQDLKE 1250
Qy 299 MTLNSTTKRNTGSPDRPKSKS-----KTDNSSLSPAPKLVCH-----VHLKKS 345
Db 1251 IEVEKEELEKISSTTSELTKKSEEVLLQEQINEQGLEIQNLKAAASHEAKAHTAQKQ 1310
Qy 346 LSGSPKLVKQNSKNSKSPBEHLEEMKMSPNKLVNTHFIPIKKGPPPAK----- 393
Db 1311 LESSQLKADLHKLTPQELTQKHVGQKEEVSYLVGQGEKEQTITVOTEMEEQE 1370
Qy 394 ----PGKSHDKPLKAKGRSKGILNGQKSTGN-SKSPKPKGLTKPTKMKMQLTLDMAKGTQK 449
Db 1371 RLKALHTQLEWQAKEHEERLAQVVEICELKQPKPELEESKAKQQLKQAALISRK 1430
Qy 450 MTRAPNSGCTPRTSSKPH-KHL-----PPAHLHLIAYYKENKDR--- 488
Db 1431 EALKENKSLQELSSARDAAVEHLTKSLADVESQVSVQNOEKDALLGLKALLOBERDKLIV 1490
Qy 489 -----EDKRSALSCVTSKTLARLLSSEDRARLPPELRSL-----VOK 524
Db 1491 EMDKSLLENQSLGGSCESLKLALGGITDEKEMKELESVRCSTAESTEWEKHKELQK 1550
Qy 525 RYELL-----EHKRW-----SMSEQRKEYLKKREELKKLKEKAKERREKEMLE 572
Db 1551 EYEVLLQSVENVSNEAERIQHVSVSRQEQVYAKLSAESDKREREKQLQDAEQEMEE 1610
Qy 573 RLEKOKRYEDQ-----ELTGKVLPAFRLVDTPEGLPN-----TLPGDVAMVVEFLSCYSG 622
Db 1611 MKEKMRKFAKSKQOKILEEEN-DRLRAEAQPVGGANESMEALLSSNASLKEELERIT- 1668
Qy 623 LLLPDAQYPIITAVSLMEALSADKGGFLYLNRLVLI-----LLQTLQLTL----- 666
Db 1669 -----LEYK-TLSKEFEALMAEKNTLSEETRNKLQVEAQELQKASLETTEKSDPKDVI 1722
Qy 667 -----LQDEIAEDYGELGWLKSEIPLTL-----HSVSELV----- 696
Db 1723 BEVTEAVVCKSQEQDSLEN-----AKLEDAEATLLANSAPGVSETPSSHDDINNLYQ 1777
Qy 697 -----RLCLRRSDVQEESESDTDNDKDSAAFEDNEVQDEFLEKLE----- 737
Db 1778 LDQLKGRITAELEMEKQKRELSTOLENEKNALLTQISAKDSBELKLEEEVAKINMLNQOI 1837
Qy 738 -----TSEFFELTSEKQLILTALCHRI-----LMTYSVQDHMETR 773
Db 1838 QEELSRVTKLKTAESEKDDLEERLMNLQALENGSIIGNYYQDVDTAQIKNEQSEMQNL 1897
Qy 774 QQMSAELWKERLAVLKEENDKKRAEKQKRB-----MEAKN 809
Db 1898 KRCVSELEEEKQLVKE---KTKVSEIRKEYMEIKQAQKPGSKIHAKELQELLKEQ 1954
Qy 810 KENGKVENG-----LQKTRKKGRIIVKPEQVDTAEADMISAVK----- 847
```


Db 1955 QEVKQLQXCIRYLGRISALEKTVKALFVHTESQKDLDTAKGNLAQAVEHHKKAQELS 2014
QY 848 -----SRLLAIQAKKERIEIQREMKVKLE-ROABEERIRKHAAAEK----- 889
Db 2015 SPKILLDDTQSAARVLADNLKXKELQSNKESIKSQIKQKDEDLRLLEQAEEKHREK 2074
QY 890 -----AFQEGITAKAKLVNRRPTIGTDRNHNRYWFLPSDEVPGLF 927
Db 2075 KMQEKLDAHLHREKARHVEDTLAEIQVSLTRKDKMKELQOOSLDSTLAQLAFTKMSLSQ 2134
QY 928 IEKGWVHDSI---DYRFNH--HCKDHTVSGDDYCPRSKKNALGNKSNNTQHTGATEVA 982
Db 2135 DRRDRVIDEAKWQRFQDAIQTKEEVRLKEENTALK-----DQLQWTH--WSELK 2187
QY 983 VETTPKQOQNLW-----FLDCSQKELDBLL 1008
Db 2188 ITVSLRLEHDKETWESKAOTELHQOKAYDKLQENKELMSQLEAGQLYHDSKNELTKE 2247
QY 1009 NCLHPQIGRESQJERLEKRYODIITHSLARKPNLGLKSCDQOELLNFLRSLDI---E 1065
Db 2248 SELSKLQOQSTDLKNSLEKREHNNLEGIITKQBEADIONKFNCEQ---LETDLTASRE 2304
QY 1066 VATRLQKGLGVVETSFEARVISLEKLFKQFCEVIALQASVIKKFLQGFMAPKQKRRK 1125
Db 2305 LTRHLH-----DEINVEQKLIISLSGKE-----EALQVAIAELHQHSEIKELN 2352
QY 1126 LOSDSAKTEVEDEBKWVEEA-KVASALEKWKTAIRAQTPSRMHVLLGMLDACIKWDM 1184
Db 2353 LSQEEENLTLBEENKRAVEKTNQLTEATETIKESLEQK-----AQLDSPVKSMS 2403
QY 1185 SAENARCKVCKGDEDDKILDCENKAPHLCLPALYVEPDGQWQCPACOPATARN 1244
Db 2404 SLQDDR-----DRIV-----S 2414
QY 1245 RGNVTERSASDSDSDSEDEEBEEDVEYAGLRRLRPRKTIRGKHSVIPPAAR 1304
Db 2415 DYQLEERHLSVILEKDELIOQAENNKLE---EIRGLR----- 2452
QY 1305 SGRRFGKPHSTRSQKAPPVDDAEVDEL-----VLQTKRSSRRQSLE--LQKCEIL 1356
Db 2453 -GHMDDLSENAKL-----DAELIQYRRDLNEVITIKDSQKQLLEAQLQNKELR 2502
QY 1357 HKLVKYSRWPPEPTREADEYDVITHPMDFTQVKNKSCGYSRVSQBELTDMQV 1416
Db 2503 NECVK-----LEGRKLGSEAEK--OSLQMSLDALOEENQGLSEIKSFKEOLTALH--- 2551
QY 1417 TNAEVNCRGSHVLSQWVKTEOCLVLLHLKHPGHYPYRRKRKFPDRLADEGDSPEA 1476
Db 2552 -----ESGALAVYHAQL-----RVRESEVQKLTAA 2576
QY 1477 VQQRDEDRRSREAIQEWL-----QDTSLYSAKINSK-----DHCNFMMLVNTQ 1521
Db 2577 LSSS-----QKRTVDLQELVCVQKEASKVSEIEDKLKRELKHLHNAIGMRNETE 2628

RESULT 15
A45592
C:Species: Plasmodium falciparum
C:Date: 22-Nov-1993 #sequence revision 02-Dec-1994 #ext change 09-Jul-2004
C:Accession: S24597; A45592; S29393; S34842; B45592; C45592; D45592
R:Zhu, J.; Hollingdale, M.
submitted to the EMBL Data Library, November 1990
A:Reference number: S24597
A:Accession: S24597
A:Molecule type: DNA
A:Residues: 1-1909 <ZHU>
A:Cross-references: UNIPROT:Q25893; EMBL:X56203; NID:g9915; PID:g9916
R:Zhu, J.; Hollingdale, M.R.
Mol. Biochem. Parasitol. 48, 223-226, 1991
A:Title: Structure of Plasmodium falciparum liver stage antigen-1.
A:Reference number: A45592; MUID:92107224; PMID:1840628
A:Accession: A45592

A:Molecule type: DNA
A:Residues: 1-195638-688;1165-1215;1590-1909 <ZH2>
A:Note: sequence extracted from NCBI Backbone (NCBIN:75010, NCBIN:75012, NCBIN:75014, NC
R:Guerin-Marchand, C.; Druilhe, P.; Galey, B.; Londono, A.; Patarapotikul, J.; Beaudoin,
Nature 329, 164-167, 1987
A:Title: A liver-stage-specific antigen of Plasmodium falciparum characterized by gene c
A:Reference number: S29393; MUID:87315391; PMID:3306406
A:Accession: S29393
A:Status: preliminary
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 323-387 <GUE1>
A:Cross-references: EMBL:M28266
R:Guerin-Marchand, C.; Druilhe, P.; Galey, B.; Londono, A.; Patarapotikul, J.; Beaudoin,
submitted to the EMBL Data Library, April 1992
A:Description: a liver-stage-sepcific antigen of plasmodium falciparum characterized by
A:Reference number: S34842
A:Accession: S34842
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 323-381, 'HKAI' <GUE2>
A:Cross-references: EMBL:M28266
A:Note: difference at carboxyl end due to frameshift error
C:Comment: This protein is found as flocculent material in the parasitophorous vacuole.
C:Superfamily: trichohyalin; calmodulin repeat homology
C:Keywords: EF hand
F154-1629/Region: 17-residue repeats (A-K-E-K-L-Q-E-Q-S-D-L-E-Q-E-R-R)

Query Match 3.7%; Score 294; DB 2; Length 1909;
Best Local Similarity 19.7%; Pred. No. 1.5e-05;
Matches 304; Conservative 273; Mismatches 584; Indels 384; Gaps 65;

QY 36 EYEARLER---YSEIRWTCKSTGSSQLTHKEAWEEO---EVAELLKEEPPAW-----Y 84
Db 533 QEQOSDLEQLAKEKLEQEQOSDLEQLAKEKLEQEQOSDLEQLAKEKLEQEQOSDLEQ 592
QY 85 EKLVLVEMVHHTASLEKLVDTAWLEIMTKYAV-GECDPFEVGEKMLVKYIVK---IHP 139
Db 593 ERLAKEKLEQOSDLEQ-----ERLAKLEQOSDLE--QLAKEKLEQEQOSDLEQ 643
QY 140 LEKVDEEATEKSGACDSPSDKENSQIAQDHQKKEVVKVEDGREGSINDRARRSPR 199
Db 644 ERLAKEKLEQOSD--LERTKASKE-TLQEQOSDLEQLAKEKLEQEQOSDLEQERRAKE 700
QY 200 KLPTSLKGERKWAPPKPLPHKYDVKLQNEDKIISNVPADSLIRTERPNKEIVFYR 259
Db 701 KLEQEQOSDLEQ-----RRAKEKLEQ-----QSDLEQERRAKEKL----- 736
QY 260 NALRAGTCENAPVVEDELVKYSLPSKFSDFLLDPYKMTLNPSTKRNKTCSPDRKPSK 319
Db 737 -----QEQOSDLEQERRAKEKLEQEQOSDLEQLAKEKLEQ-----EQOSDLEQERRAKE 785
QY 320 KSKTDNSLSPLNPKLWCHVHLKSLSGSPLKVNKS--KNSKSPHEHLEEMKMWSPNK 377
Db 786 KLEQEQOSDLE-----QDLAKEKLEQEQOSDLEQERRAKEKLEQEQOSDLEQERR 832
QY 378 LHTNFIHPKGPAPKPKGHSKDPKAKGRSGKILNGQKSTGNSKSPKGLTKPTKMQK 437
Db 833 L-----AKEKLEQEQOSDLEQERRAKEKLEQEQO-----SDLEQDLAKEKLEQEQO 876
QY 438 MTLDMAGTKQMTRAPRNSGGTPTSSKPHKHLPPAALHLIAYYKKNKDRKEDKESALS 497
Db 877 SDLEQERRAKEKLEQEQOSDLEQERRAKEK-----LQEQOSDLEQERLA--- 919
QY 498 VISKTARLLSSEDRARLPFEELRSIVQKRYELLEHHKRWASMESEOR-KEYLKKKKREELK 556
Db 920 -----KEKLEQEQODLEQERRAKEKLEQEQOSDLEQERRAKEKLEQEQOSDLEQ 966
QY 557 K--LKEAKERERKEKMLELKEKQYEDQ-ELTGNLPAFLVDTPEGLPNTLFGDVANV 613
Db 967 ERLAKEKLEQEQOSDLEQLAKEKLEQEQOSDLEQLAKEKLEQ-----QEQOS----- 1013
QY 614 VEFLSYSGLLLPDAQYPTITAVSLMEALSAKGGFLYNRLVILVILLQTLQTLQDEIAE 673

Db 1014 -----DLEQLRAKEKLOQOQSD-----LEQLRAKEKLOEQSDLEQLRAK 1056

Qy 674 DYGLGKMLSEIPLTHSVSELVRLCL-----RRSDVQBS-----EGSDTDNDKDSAA 722

Db 1057 E-----KLOE-----QQSDLEQLRAKEKLOQOQSDLEQLRAKEKLOQOQSDLEQLRA 1106

Qy 723 FEDNEVDFLEKLETFSEFELTSEKLOILTALCHRLMTYTSVQDIMEYR-----QQ 775

Db 1107 KEKLOQOQSDLEQ-----ERLAKEKLO-----QQSDLEQLRAKEKLOE 1146

Qy 776 MSALMKERLAVLKEENDKKRAEKQKKEKMEAKNKEN-GKVENGKGTDRKKRIVKFEPO 834

Db 1147 QQSDLEQLRA-----KEKLOEQSDLEQLRAKEKLOEQSDLEQLRAKEKLOEQSDLEQLRA 1199

Qy 835 VTEAEDMISAVKSRRLATQAKKERRIQERMKVKLERQ---AEEERIRKHAAAEKA- 890

Db 1200 SDLEQLRAKE-----KLOEQSDLEQLRAKEKLOEQSDLEQLRAKEKLOEQSD 1252

Qy 891 -FOEGTAKAKLVMRPTPIGTDNRHNRWYLFSDDEVFGLFIEKGWVHDSIDYRFNHCKDHT 949

Db 1253 LEQLRAKEKLOEQSDLEQER-----RAKEKLOEQ 1284

Qy 950 VSGDEYCPBSKCANLGKSNAMTQHGCTATEVAVETTPKQGNLWFLCDSQKELDELLN 1009

Db 1285 SDLEQER---RAKEKLOEQSDLEQLRAKEKLOEQSDLEQERL-----AKEKLOEQSD 1336

Qy 1010 CLHPOGIRESQLKERLEKRYQDIISHILARKFNILGKSDGNQ-----ELNFLASDLI 1064

Db 1337 DLE---QERRAKEKLOEQSD-LEQLRA-KEKLOEQSDLEQLRAKEKLOEQSDLE 1390

Qy 1065 EVATRLKGGGLGVVETSEPEARVISLEKUKDQFGEVIALQAVIKKFLQGFVAPKQKR 1124

Db 1391 Q--DRLAKEKLOEQORDLEQERR--AKEKLOE-QQSDLEQLRAKEKLOEQSDLEQERR 1445

Qy 1125 ---KLOESDSAKTEEVDEEKMKVEAKVASALEKWKTAIREATQFMRHVLGLMDACIK 1181

Db 1446 AKEKLOEQSDLEQLRAKEKLOEQ---QSDLEQLRAKEKLOEQOR----- 1489

Qy 1182 WMSAENARKVCPKKGEDDKILDCENKAFHLCRLPALYVVPDGEWCPACOPATAR 1241

Db 1490 -LEQER-RAK---EKLOEQSDLEQ-----RAKEKLOEQSDLEQLRAKEK 1534

Qy 1242 RNSGRNYTEE-SASDSEDDSDSEEEEEEEEEEEEDVEVAGRLRPRKTRKGHSVIP 1300

Db 1535 LQEQORDLEQLRAKEKLOEQSDLEQLRAKEKLOEQSDLEQLRAKEKLOEQSDLE 1594

Qy 1301 PAARSGRRPKKPHSTRSQPKAPPVDDAEVDLVLQTKSSRRQSLELQKCE-EILHKI 1359

Db 1595 QERLAKEKLOEQORDLEQLRAKEKLOEQORD---LEQRKADTKKRLKKEHGDVLAED 1651

Qy 1360 VKYRFSWPFREPVTRDEADYDVITH---PMDFQTVQNKSCGSYRSVQEP---LTD 1411

Db 1652 LYGRLEIPALELPSENE-RGYI--IPHQSLPD-----NRGNSRDSKEISIIETKN 1700

Qy 1412 MKQVFTNAEVYNGRSHVLSCMWKTEQCLVLLHKHLPGHYVRRKRPORLBADEGD 1471

Db 1701 RESITTNVE--GRDDIH-----KGH 1718

Qy 1472 SEPEAVQQRDEDRRREAEIQEWLQDTSIYSAKINSKDHNCNFM 1516

Db 1719 LEEKDGSIRPEQEDKSDIQNHHTLET-----VNISVNDPQI 1757

RESULT 16

T16270
hypothetical protein F35D11.11 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 09-Jul-2004
C:Accession: T16270
R:Fulton, B.
submitted to the EMBL Data Library, June 1995
A:Description: The sequence of C. elegans cosmid F35D11.
A:Reference number: Z18487

A;Accession: T16270
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1827 <FUL>
A;Cross-references: UNIPROT:Q20042; EMBL:U29381; NID:g868214; PID:g868224; PIDN:AAA6875
A;Experimental source: strain Bristol N2
C:Genetics:
A;Gene: CSPF:F35D11.11
A;Introns: 76/2; 131/3; 159/3; 185/3; 221/3; 253/3; 320/1; 869/3; 1133/3; 1205/2; 1250/

Query Match 3.6%; Score 290; DB 2; Length 1827;
Best Local Similarity 20.1%; Pred. No. 2.2e-05;
Matches 317; Conservative 262; Mismatches 626; Indels 370; Gaps 73;

Qy 121 DFEVGEKEM-----LKVKIVKIHPLEKVEDEATEKKSDGACDPSSS---DKEN-SSQIA 170

Db 55 DIGVGEENLDELPAKLKESLFKNNTLEEVENMLRE-ENDAALAANEHLRVDTATNLSRQLQ 113

Qy 171 QHQKKETVVKDEGRRESINDRARRSPKLP-----TSUKGERKWAPP 215

Db 114 QLQOQOHT-----ESMRFRSENTSRELQYCKGPEICNTFLFLFSCKPYQERYNQ 164

Qy 216 KFLPHKYDVKLQNEDKIIS-----NVPADSLIR--TERPPNKEIVRPIRHNALRAGT 266

Db 165 TETQHRKGLISLWKEFTAVKRLQHLRTTTTANDLDRQLTFTTCATLMRKAIRH----- 217

Qy 267 GENAPMVVEDELVKYKSLPSKPSDFL---LDPYKVMTLN--PSTKKNKGTSPPDK---P 317

Db 218 -----AEQNLDDQEKMKREKDDVLDLRLQLNSVTENYMKSEKANEQRDLKKEDE 271

Qy 318 SKKSKTDNSLSPLN--PKLMCHVHLKKSLSGSPKLVKNKSNKS-----PEHLBEMM 370

Db 272 CRKLREONDELSDILEQLSKMAHAGGRSQSPSVDVNTTTHAFLNFTETPMDVARQMR 331

Qy 371 KMSPNKLTHTNPHIPKGGPPAKPKGKHSKPL-----KAKGRSKGIUNGKSTGNSKSPKK 426

Db 332 KLLTTK---NGEIDSEAREAKQAEKERDRAKKDLKEEKKRRKDDREAREKSSVVSQRE 387

Qy 427 -GLK-----TPTKMKQMTLDMAGTKQMTAPRNSGGTPRTSSKPHHLPALALHIA 480

Db 388 HDLKKLDDLRAKASEKIRNLEQRSQEKLTISVQNS---LNEAHRQHQFIEEL--MIR 442

Qy 481 YYKENKQDREKRSALSCVISKTARLLSSDRARLPEEL--RSLVQKRYELLEKKR--- 534

Db 443 HREELKEREDSH-----EALRSKQTEERSRFEKERSEREKIRRESDELRETQSLKG 495

Qy 535 -WASWSEE-----QRKEYLKKRBEELKKLKEKAK--ERREKEMLE-RLEKQKRYEDQEL 585

Db 496 DVAAKMTDLDDKTLRLDMLETERDELKKKLETEREQADQORDLEIAECRAKLDEMAEKEAE 555

Qy 586 TCKNLPAPRLVDTP---EGLPNTLFGDVAMVVEFLSCYSGLLLPDAQYPIITAVSLMEALS 642

Db 556 LKELAEAFQAIITAMEGEGKLNQ-----EQFLESKNEL-----NTLTDQIESLN 599

Qy 643 ADKGGLYLNRLVILLOTLLQLQDEIT-----AEDYGLGMLKLEIP 686

Db 600 SEVEN---KNEE---IRNLMATLQEKVHIQNVRTSSHQLTATYEEANGEIDILKAEIT 652

Qy 687 LTLHSVELVRLCLRRSDVQBSSESDTDNDKDSAAFEDNEVQDEPLE--KLETSEFFEL 744

Db 653 RLHEQVNE-----RTRQISEANEKYDDAARKNDALLEDVATWQEKYEQLKMLEBEMNR 706

Qy 745 TSEEKLO--ILTALCHRLMTYTSVQDHMETROQMSAELWKEKRLAVLKEENDKKRAEKQR 802

Db 707 GQEKEREADLRALLDDLRGNFDKUTNELKQGVTVDSLNEEISLSLQOLNK--SEKERK 764

Qy 803 KEM-----EAKNKENGKVENGLGKTRDKKRIKIVKPEQVDTEAEDMISAVKSRRL 852

Db 765 BELLRWEELEQKNEAEEMKEEYEVKLQAEKOR-QGVENFGKECEARMNEL---TKIHEML 820

Qy 853 A-----IQAKKEREIQEREMKVLER-----QAEEERIRKHAAAAEKA 890

Db 821 MEEHDQLKVDHLHTBEEVERLKEKMRKELEKLNQDGDRAEWSNERNRLESSKNEAVTE 880

QY 891 FOEGIAKALVMRRTPIGTDRNHNRYWLPFSDVPGLFIEKGWVHDSIDYRFNHCKDHTV 950
DB 881 LQERVOKLEDDVVK-----EKDEKETAIRRD-----LEDS-----HEK 912
QY 951 SGD-EDYCPRSKAMLGKASNMNTQGTATEVAVETTPPKQGNLWFLCDSQKELDELIN 1009
DB 913 SRDLDDKLARKMELTDEEKEDRKQKTLNEERMKLM-----EQKEEAMLVA 959
QY 1010 CLHPQGIRESQKLERLEKRYQDIHSIHILARKPNGLKSCDQGNQELLNPLRDLLEVATR 1069
DB 960 TKHAHTI--DQOTRISVLEGDV-----EKLTAGIAE--RESSINALENTMELISK 1007
QY 1070 LQKGGLYVEETSEPEARVISLEKLPFGCEVIAL-QASVIKKFQGFMAP-KQKRRKLO 1127
DB 1008 L-----ETFEAB-----LEKUD--ELAVMLKQNSLXNGKEGLSEKWNBERKKIQ 1051
QY 1128 SEDSAKTEVDEBEKQVBEAKVASA-LEKWKTAIREAQT--FSRMHVL-LGMLDACIKWD 1183
DB 1052 D-----LADQLREANKVVMRMKQVNLLEKKKQELDQNTDITNKVRQLEIQMLDKAAKVE 1107
QY 1184 MSAENARCKVCPKGEDDKLILCDBCNKAFHFLCLRPALYVDPDGEWQCPACQATARRN 1243
DB 1108 VSGDLLR-----KMEHDAQMLKQAGNEQFRLTDLEKVRKALQD-ENQRLVNDLATVRAA 1161
QY 1244 SGRNYTEESASEDEDDSEDEEBEEEEEE-----DYEVAGLRRLRPRTIRGK---- 1295
DB 1162 FEVKGRETSKASISDLTKYRABEAKNGELDNQRLSLATVTLKLE-ROSLKAKDSN 1220
QY 1296 -----HSVIPPAAARGRRPGKPPHSTRSQPKAPPVD-----DAEVDE 1333
DB 1221 RLDSQKRFEEVQSKLANLOKSAVESLQNPMSNQRNSIYVDIPRAASSITGLNENSD 1280
QY 1334 LVLQTKRSR--RQSELEKQCEIILHKI-----VKYRFSWPFPPVTRDEAE--DYVDVIT 1385
DB 1281 VPLRSPSVRFADSSQNMQRVSDSVSGVTLLRF--LKERIEQLEADNADLSDALE 1337
QY 1386 HPMDFTQVONKSCG-----SYRSVQEFLLDMQVFP-TNAEVNCRGS 1427
DB 1338 KAKDELQRNEKLADQMWIERVQLVHITERNTIENRMTSQRMVLTNEESSRSRHH 1397
QY 1428 HVLSCMVKTEQCLVLLKHLPHPYVRRKRFDPDLAE-----DEGDSPEAVG 1478
DB 1398 ETRSMKAR-----ISTLEHL-----REKESKLHLRKEIEVHLQHLDALSESKEATG 1446
QY 1479 QSRDEDRRSREAEIQ 1493
DB 1447 LVGVQDSKHDLEEQ 1461

RESULT 17

T13564
microtubule-associated protein homolog - fruit fly (Drosophila melanogaster)
N/Alternate names: hypothetical protein EG:4984.1
C/Species: Drosophila melanogaster
C/Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 09-Jul-2004
C/Accession: T13564
R/Spinos, L.; Papagiannakis, G.; Siden-Kiamos, I.; Louis, C.
submitted to the EMBL Data Library, April 1999
A/Description: Sequencing the distal X chromosome of Drosophila melanogaster.
A/Reference number: Z17689
A/Accession: T13564
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-5327 <SPA>
A/Cross-references: UNIPROT:O76891; EMBL:AL031128; PIDN:CAA20006.1
C/Genetics:
A/Cross-references: FlyBase:FBgn0025392
A/Introns: 24/2; 52/3; 104/3; 179/1; 232/1; 1669/3; 2566/1; 4798/3; 5272/1
A/Note: EG:4984.1
C/Superfamily: Drosophila 576K microtubule-associated protein homolog
Query Match 3.6%; Score 288.5; DB 2; Length 5327;

Best Local Similarity 20.1%; Pred. No. 8.9e-05;
Matches 274; Conservative 213; Mismatches 491; Indels 385; Gaps 62;
QY 115 AVGBECDVEQEKERMLKVIVKIHPLEKVDDEATEKSGDAGCDSPSSSKENSSQIAQDHQ 174
DB 2685 SVFESVKDEHDKAESRRRESIAKV--ESVIDEA-----GKSDSKSS-----SODSQ 2727
QY 175 K--KETVYKDEGBRESI-----NDPARESPKLPTSLAKGGERKWAAPPKFLPHKYDVKLQN 228
DB 2728 KDEKSTLASKASRRSVSVESKDDAEKSRPESVIASGEP-----VPRESKPLDS 2780
QY 229 ED-----KIIINVPADSLIRTERPNKEIVIRYFIRHNALRAGTGENA--PWVVEDELVK- 280
DB 2781 KDTSPGVSVESVTAED--EKSEQSRRESVAESVKADTKDGKSQEAESPSSV-DELLKD 2838
QY 281 -----KYSL-----PSKFSDFLLDPYKYMTLNPSTTKRNT-----G 311
DB 2839 DDEKESRRQISITGSHKAMTWGDESPMDKADKKEPSRPESVAESI KHENTKDEESPLG 2898
QY 312 SPDRKPSKSKTD-NSSLSSPLNPKLWCHVHLKSLGSLPKVKNSKNSKSPHEHLEMM 370
DB 2899 SRDSVAESI KSDITKGEKSPLPK--EVSRRPESVWGS---IKDEK-AESRRRESVAESV 2951
QY 371 KMSPNKLHTNFHIPKGGPPAKPGKHSKDKPLKAKGRSGKILNGOKSTGNSKSPKGLKT 430
DB 2952 KPES-----SKDATSAPSKESHRPESVLSGLD--EGDKTTSRRSVSVADSIDK 2998
QY 431 PKT-----KMKQMTLLDMAGTKQMTAPR-----NSGGTPRTSSPKHKLPPAA 475
DB 2999 EKSLIVSQEASRPSEAESEALMDAAAPQETSRPESVTSVKDGKSPVASKASR---PAS 3055
QY 476 LHLIAYYKENDREDKRSALSCVLSKTARL-----LSSDDRARLPER--LRSLVOKRYE 527
DB 3056 VAENA--KSDADESKQRPESLPDSKAGSKIDEKSPASKOEAESKESRRRESVAEQFP 3113
QY 528 LLEHK-KWASMSBQRKEVYLKKRE-----ELKKLKEKAKERREKEMLERL 574
DB 3114 LVSEKVPSPASVAESVKDEAEKSEESPLMSKEASRPASVAGSVKDEAEKSEESRRSV 3173
QY 575 EKQRYEDELTKGNLPAPRVLVDPTEGLPNTLFGDVAMVVFPLSCYSG-----LLLPDAQ 629
DB 3174 AEKSPSPKE-----ASRPASVAESVKDEADKSEESRRRESGAESKSPASKAS 3222
QY 630 YPIT-AVSLMEALSADKGPVLYLNRLVILLQTLLOLDEIAEDYDYGELGKMLSEIPLT 688
DB 3223 RPASVAESI KDEAEKSK-----EESRRRESVAB-----KSPDP 3254
QY 689 LHSVSELVRLCLIRRSDDVOESEGSDTDNDKDSAAFP-----NEVODEFLE 734
DB 3255 SKEASRPTSVA--KSVKDEAEKSEESSRDSVAESKSPASKASRPASVAESVQDE-AE 3310
QY 735 KLETSEFFELTSEKQLQILTALCHRLMTYSVQDHME--TROOMSLELWKLAVLKEEND 793
DB 3311 KSKEESRRRESVAESKPLAYKEASRPASVAESI KDEAEKSEESRRRESVAESKSPASKAS 3370
QY 794 KGBAEKQKREKWEAKNGKNGKVENGLGKTDRKKRIVKPEPQVDTEADMISAVKSRLLLA 853
DB 3371 RPTSVAESVKDEAEKSEES-----RDSVAESKSPASKASRPASVAES---VQ 3417
QY 854 IQAKKERIQEREMKVKLERQAEERIRKHKAA-----AEKAFQEGIAKAKLVNRRPTPI 907
DB 3418 DEAEKSEESRRRESVAESKSPASKASRPASVAESVKDEAEKSEK--SRRESVAESKSP 3475
QY 908 GTDRNHNRYWLPFSDVPGLFIEKGWVHDSIDYRFNHCKDHTVSGDEDYCPRS--KKANL 965
DB 3476 AS-KEASRPASVAESV-----KDEAEKSEESRRRESVAESKSP 3512
QY 966 -GKASMTQGTATEVAVETTPPKQGNLWFLCDSQKELDELIN--CLHPQGIRESQK 1022
DB 3513 PSKEASRPT--SVAESVKDEAEKSEESRR-----ESVAESKSLASKASRPASVAES-VK 3565
QY 1023 ERLEKRYQDIHSIHILARKPNILGLKSCDGNQELLNPLRSLDLEIATRLQKGLGYVEETS 1082

Db 3566 DEAEKKEE-----SRRESVAEKSPLASKE-----ASRPAS-----VAESV 3601
 Qy 1083 EPARVISLEKLDPGCEVIALQASVKKFLOGFMAPKQRRKLOSEDSAKTEBVEDEKK 1142
 Db 3602 KDEA-----EKSKE-----VSRESVAEK-----SPLPSKASRPTSVAESVK-----DEADK 3644
 Qy 1143 MVEEAKVASALEKWKTAIRAEQTFSRMHVLLGMLDCAIKWDMGAENARCKVCPKKGEDDK 1202
 Db 3645 SKEESRESGAESPLASMEASRPTSV-----AESVK-----3676
 Qy 1203 LILDECNKAFLFLCLRPALYEPDGEWQCPACQATARNRSGRNVTES-----ASE 1256
 Db 3677 -----DETEK-----SKEESRESVTEKSPLSKEASR 3704
 Qy 1257 DSEDDSEDEEEEBEEDYEAGLRPRKRTIRGHSHVTPPAARSGRRP-----1309
 Db 3705 PTVSAESVDEAEKSEES-----RESVAEKS-----PLAKSESSRPASVAESI 3749
 Qy 1310 -----GKKPHSTRSOPKAPPVDDAEVDLVLQTKSSRRQSL 1347
 Db 3750 KDEAEGTKOESRESPESGKAESIKGDQSSLASKETSRPDSV 3792

RESULT 18
 JC5421
 Smooth muscle myosin heavy chain 2 - mouse
 C:Species: Mus musculus (house mouse)
 C:Date: 10-Jun-1997 #sequence_revision 18-Jul-1997 #text_change 09-Jul-2004
 C:Accession: JC5421
 R:Hasagawa, K.; Arakawa, E.; Oda, S.; Matsuda, Y.
 Biochem. Biophys. Res. Commun. 232, 313-316, 1997
 A:Title: Molecular cloning and expression of murine smooth muscle myosin heavy chains.
 A:Reference number: JC5420; MID:97242182; PMID:9125171
 A:Accession: JC5421
 A:Molecule type: mRNA
 A:Residues: 1-1938 <HAP>
 A:Cross-references: UNIPROT:Q08639; DDBJ:D85924; NID:g1945079; PID:BAA19691.1; PID:g1945079
 A:Experimental source: smooth muscle
 C:Comment: This protein plays a role in smooth muscle cell contraction.
 C:Superfamily: myosin heavy chain; myosin motor domain homology
 C:Keywords: nucleotide binding; P-loop
 F:88-771/Domain: myosin motor domain homology <MWOT>
 F:178-185/Region: nucleotide-binding motif A (P-loop)

Query Match 3.6%; Score 287.5; DB 2; Length 1938;
 Best Local Similarity 19.9%; Pred. No. 3e-05;
 Matches 259; Conservative 224; Mismatches 499; Indels 319; Gaps 57;

Qy 12 LVNPLGEEPPFTPIPTQSAFRT-----REEYAEARLYSERIWTCKSTGSSQLTHKEA 65
 Db 750 MIKALELDPNLVRIQSKIFFRTGVLHLEERDLKI---TDVIMAFQAMCRGYLARKAF 806
 Qy 66 WEEEOEVAELLKEEFPANWEKLVLEMHVHTASLEKLVDTANLEIMTK-----YAVGEBC 120
 Db 807 TKRQQLT-----AMKVIQRNCAAYLKLNRQWMLRFTKVKPLLVQVTRQEE 852
 Qy 121 DFEVGKKEKMLKV-----KIVKHPLEKVDDEATEKKS-----DGACDPSGSDKENSQ 168
 Db 853 EMQAKEEMQKITEROQKAETELKEHQTLQAEKTLLOEQLOAETELYAESEMRVR 912
 Qy 169 IQADHQKKTIVKVEDGRRRESINDRARRSPRLFTSLKGGERRKAPKPLPHKYDVKLQN 228
 Db 913 LAAKQQLSEILHEMARLEEEEDRRQ-----LQAEK-----KMAQQMLDLEQL 959
 Qy 229 EDKIISNVADSLIRTERPNKEIVYFIHNLALRAGTCENAPVVEDELVKYSLPSKF 288
 Db 960 EE-----BEAARQKLO-----LEKVTAEAKIKKLEDDILVMDQNSKL 997
 Qy 289 S---DFLLDPYKWTLPNTPKRNKNTGSPDRKPSKSKSDNSLSLSPNPKLWCHVHLKKS 345
 Db 998 SKERKLLERVSVDLTTLNLAEEBEKA-----KNLTKLKSKHESMISEL-----EVLKKE 1046
 Qy 346 LSG-----SPLKVNKSNKSPKEEHLLEMMQMMSPNKLTHTNPHIPKGPAPKPGKHSKDP 401

RESULT 19

T18296

myosin heavy chain - Entamoeba histolytica

C:Species: Entamoeba histolytica

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004

C:Accession: T18296

R:Guillen, N.

submitted to the EMBL Data Library, February 1997

A:Reference number: Z18865

Db 1047 EKSQBLEKRLKLEGDASDPHEQIADLOAQIAELKM-----OLAKKEEELQALARLDEE 1102
 Qy 402 LKAKGHS-KGI-----LNGOKSTGNSKSPK-----GLKTPKTKMKNQMTLLD 442
 Db 1103 IQAKNNALKKIRELEGHISDLQEDLDSEARAANKAKQKRDLDGEELAEUKTELED---TLD 1160
 Qy 443 MAGTKQMTTRAPNSGGT-----PRTSSKPHKLPPAALHLIAYIKENKOREDKRSAL 495
 Db 1161 -STATQOELRAKREQEVTVLKKALDEBETRSHEAQVQEMQKHTQAVBEELTEQLEQFKRA- 1218
 Qy 496 SCVISKTARLLSSEDRARLPBELSLVQRYELLEHHKGRWASMSBEQRKEYLKKGREELK 555
 Db 1219 KANLDKSKOTLEKEN-ADLAGELRVLQAKQB-VEHKKK-----KLEVOLQDLQ 1265
 Qy 556 KKLKBAKAKERREKEMLERLEKOKRYVEDQELTGKNLPAFLVDTPEGLPNTLFGDVAWVVE 615
 Db 1266 SKGSD--GERARAEISDKVHKLQN-EVESVTG-----MLNEAEGKAIKLAKDVASLGS 1315
 Qy 616 FLSCYSGLLLPDAQYPITAVSLMEALSADKGGFLVNLRVLLVILLQTLQTLQDEIAED- 674
 Db 1316 QLQDTQELQOEBETROKLVNSTKLRQLEDERS-----LQQLQDEEM 1356
 Qy 675 -----YGEIWMKLSIPLTLHSVSELVRLCLR-RSDVQSESESGSDTDNDKDSAAFE 724
 Db 1357 EAKONLERHVSTLNQLSDSKKKLQDFASTIEVMBEKGKRLQKEMEGLSQQYEEKAAAYD 1416
 Qy 725 -----DNEVQDFE-----LEKLETSEPFELTSEELQIILTALCHRLMT 763
 Db 1417 KLETKYNNRLQOELDDLVVDLQNRQLVSNLEK-KOKKFDQLLAEEK-----NISK 1466
 Qy 764 YSVQDHMETRQOQMSAEL-WKERLAV-----LKEENDKKRAEKQKREKEMAKNKENG 813
 Db 1467 YA-----DERDRAEAAREKETKALSLARALEALEAKEELERTNKMKAEMEDLVSSKD 1521
 Qy 814 KVENGIGKTDKRRKRVKPEPOVD-----TEADMISAVKSRRL-----LAIQAKKERE 861
 Db 1522 DVGKKNVHELEKSKRAL--ETQMEEMKTQLEESDDVQATEDAKLRLEVMNQALKGQFERD 1579
 Qy 862 IOER-----EMVKVLEKQ-----AEERIRKHAABAKAFQEGIAKAKLVMRRTPI-G 908
 Db 1580 LQARDEQNEEKRRQLQRLQHEYTELEDERKQRLAAAKKLEGLDKOLELQADSATKG 1639
 Qy 909 TDNRNRYWLFSDVEPGLFIEKGWVHDSIDYFNHCKDHTVSGDEDYC---PRSKKANL 965
 Db 1640 RE-----EAIKQLRKLQAOQMKD-----FQRELDADARASDEIFATSKENEEKAKS 1684
 Qy 966 GKNASMNTQHGAT-----EVAVTTTTPKQGNLWFLCDSQKELDELLNCLH 1012
 Db 1685 LEADLMQLQEDLAAARARKQADLEKELEAEELASSLSGRNT--LQDEKRRLE-----1735
 Qy 1013 POGIRESOLKERLEKRYQDIIHSIHLARKPNLGLKSCDGNQELLNFLRSLDLEVATRLQK 1072
 Db 1736 ---ARIALEEELEEEQGNMEANSRVRKATLQA-----EQLSNELEATER-STAQKNS 1785
 Qy 1073 GGLGYVEETSEPEARVISLE-----KLQDFGCVIALQASVTKKFLQGFMAPKQKR---R 1124
 Db 1786 ARQQLERQNKELRSKLQVEGAVKAKLK---STVALEAKIAQLEEQVEQEAEREKQAATK 1842
 Qy 1125 KLQSDSDAKTE---EVDSEKKNVVEAKVASALEKKNWKTARE 1162
 Db 1843 SLKQDKKKLKEVLLQVEDERKMAEQYK-EQAEGKNTKVKQ 1881

A;Accession: Tl8296
A;Status: preliminary; translated from GB/EMBL/DDBU
A;Molecule type: DNA
A;Residues: 1-2139 <GUI>
A;Cross-references: UNIPROT.Q07569; EMBL.L03534; NID:g1850912; PID:g1850913; PIDN:AAB460
C;Genetics:
A;Gene: mhca
C;Superfamily: myosin heavy chain; myosin motor domain homology
F;91-780/Domain: myosin motor domain homology <MMO>

Query Match 3.6%; Score 286.5; DB 2; Length 2139;
Best Local Similarity 18.9%; Pred.No.3.8e-05;
Matches 277; Conservative 230; Mismatches 490; Indels 469; Gaps 60;

		Qy	37	EYEARLRYSERIMTCKSTGSSQLTHKEAVEEEOEAELLKKEFPAPWYEKLVLVEMHHNT	96
Dd	:	Bb	900	ELEAKIQDLEDKI-----SEL-----ESKLSAAELDKQEL-----NLKIENLEEDK	940
Qy	:	Bb	97	ASLEKLVDTAWLEIMTKYAVGEBCDFVGK-----EKMLVKVIKHIPLEKVD	144
Dd	:	Bb	941	EELKETINLKGDGLKDSKLKGDELEVETELNSQINTLNATVDKKOKTTAEQMOSIDEKE	1000
Qy	:	Bb	145	EEATEKISD-----GACDSPSSDKENSQAQHQQK-----ETVVK	181
Dd	:	Bb	1001	DEITPKLGDIKLABEEKDDLEQRADVSATKDIDIANKKITTECDAKDEIAKLQEDEL	1060
Qy	:	Bb	182	EDEGRRESINDRARSRPKLTSLKGERKWAPPFLPHKYDVVKLOVEDKIISNVPADSL	241
Dd	:	Bb	1061	DEENKNKDLTNELQOTQLGETEKSLAAQVAATKAASDERDTLSQN-----	1107
Qy	:	Bb	242	IRTERPPNKSEIVFYRIHNALRAGCTGENAPMVVEDEVLYKYS-LPSKFSDFLDPVKYMT	300
Dd	:	Bb	1108	LENELKTNNLTN-----TKADEKKISGLQDYVEDLEDKKNIEG	1148
Qy	:	Bb	301	LNPSTKK-----NTGPSDRPKSKSYTDINSLSPLNPKLCHVHLKXLSGSPLK	352
Dd	:	Bb	1149	DLRNAQKKIKELDDIITKGADVSQLYGQKEEVESQTAKMQE-----EKGAIGNDVK	1200
Qy	:	Bb	353	VKSNSKNSPEEHLEEMKMWSPNKILHTNFHIPKGPAPAKPGKHSKP LKAKGRSGIL	412
Dd	:	Bb	1201	NK-EKTIKEVELEISOEKLDETEVEKE-----DAEKKKKEIKEKMAEQEKE--	1249
Qy	:	Bb	413	NGQKSTGNKSPPKGLTPKTMQMTLLDMAKGTQM-----TRAP	454
Dd	:	Bb	1250	NVESSNKSTEKDDKKKL-DNLKDTQKKLDMDTADNEKLKAKADLEAQLNEVDNHKAV	1308
Qy	:	Bb	455	RNSGGTPTRTSPKHKHLP PPAALHIAYK-----ENKOREDKRSALS CVISKTARLLS	507
Dd	:	Bb	1309	ADABELLNKKAQS QKELNSL KAELEALT KAASKVSVESKNKDSENEKAALESIIDQANEKLK	1368
Qy	:	Bb	508	S--EDRALRLPEELSLVQKRYE-----LLEHKKRWASMSE-EORKEYLKKREELKK	556
Dd	:	Bb	1369	NIQADLRKATADLAQANKEKAEVEAQRDKLVAUDNKWTLEIHKARDEENTYKVNEYK	1428
Qy	:	Bb	557	KLK-----EKAKERBEKEMLERLEKQRYEDOELTGKNLPAFRLVDTPEGLPNTLF GDV	610
Dd	:	Bb	1429	VLKERKEADLEEANENLDIEKKDRNMKEQV--KKLEGE-----LKETDKLN-----	1473
Qy	:	Bb	611	AMVVEFLSCYSGLLLPDAQPYPTAVSLMEALSADKGGFLYNRV-----LVILLQTLQ	664
Dd	:	Bb	1474	AATAEKDSIFTAKKQSDA-----DLEE-----LNKTVEEHDHVAVAKLANQTIT	1515
Qy	:	Bb	665	TLLQD--EIAEDYVELGMKLSIEPLTIHLSYSELV-----RLCLRSDVQ	706
Dd	:	Bb	1516	KLTRDNQSAEELNELSKADKKKISELEEQVNEUESPVTGNGADENEI KI RDAQIA	1575
Qy	:	Bb	707	E-----ESEGSTDDD---NKDSAAPBDN-----EVQDEFLEKLETS-----	739
Dd	:	Bb	1576	DLNKALEWKGVQNNLOATNKELKA-KDNDLTSKIETITENEMKKLENAKKBQDKDEAD	1634
Qy	:	Bb	740	-----EFFELTSBKLIQLTALCHRILMTYSVQDHME-----	771
Dd	:	Bb	1635	KAVSEOTIKRKGLEEBEYKLTTE-----IOALKFOINAPSVAQOESEKORBESDIAELKE	1689

F:888-2017/Domain: coiled coil #status predicted <COI>
F:888-1328/Region: S2
F:1329-2017/Region: light meromyosin
F:176/Modified site: N6,N6,N6-trimethyllysine (Lys) #status predicted
F:231/Binding site: ATP (Lys) #status predicted
F:745,755/Active site: Cys #status predicted

Query Match 3.5%; Score 280; DB 1; Length 2017;
Best Local Similarity 20.0%; Pred. No. 6.9e-05;
Matches 287; Conservative 224; Mismatches 485; Indels 442; Gaps 66;
QY 60 LTHKEAWEEQVABLLKEFPW-----YEKL-----LEMVHHTASLEKLYDT 105
DB 819 LAHLEE-ERDFKISDLI-VNFQAFRCGFLARNYQKRLQQLNAIRIIQRNCAAYLKLRW 876
QY 106 ANLEIMTK-----YAVGEECDPVGKEMKLVKIVIKHLEKVDDEATEKSDGACDSP-- 159
DB 877 QWRRLYTKVPLEVTQKEKLVQKEDLQVREKLDTLAKNTQBYERKYQQALVEKTTL 936
QY 160 -----SSDKENSQIAQDHQKKEVTVKEDEGRESINDARRSPKRLPTSLKKG 208
DB 937 ABQQAETLCAAEESRLMARKQELDMQOELTRIIEEER-----VIALGG 987
QY 209 ERKWAAPPFLPHKYDVLQN-BDKIISNVPADSLIRTERPPNKETIVRYFIHNALRAGTG 267
DB 988 EKK-----KLELNIOQLEQLEEEBAARQKLEK----- 1017
QY 268 ENAPWVVEDELVKYSLPSKSFDFLLDPKYNTLNPSTYKRNKVGSPDRKPSKSKTDNSS 327
DB 1018 -----VQDAKIKKYE-----BDLAL-----TDDQNQKLKEKKL----- 1047
QY 328 LSSPLNPKLVCHVHLKXLSGSLPKVKNKSNKSPPE-----HLEEMKMSPNKLHTNPH 383
DB 1048 LEERAN-----DLSQTLAEIEEAKAKLAKLAKHEATITIEERL----- 1087
QY 384 IPKGPAPKPKGKSDPKLAKG-RSKGILNGKSTGNSKSPKGLKTPKTKNQWTLID 442
DB 1088 -----HKDQOQROESDSK-----RKITEVADLKEQLNERRVQVDEMQ-AQ 1128
QY 443 MAG-----TQKTRAPRNGGTPTRTSSPKHPLPPAALHLIAYIKENKDRKRSALSCV 498
DB 1129 LAKREBELTQTLLRIDEES-ATKATAQAKRELESQ-----LAEIQEDLEAE----- 1174
QY 499 ISKTARLLSSEDRARLPEELRSIVQKRYELLEHKKRWASMSSEQKRYLKKREE-----L 554
DB 1175 --KAARAKAEKVRRLSELEAL-----KNELLD-----SLDTTAAQOELRSKREQLATL 1223
QY 555 KKKLKEKAKERKEKEMLEKRYEQELTKGNLPAPRLVDTPEGLPNTLFGDVAMV 614
DB 1224 KKSLEETVNHGV-----LADMRHKGSELNSIN-----DQLENLRK-----AKTVL 1266
QY 615 EFLSCYSGLLLPDAQYPTAVSLMALSADKGGFLYNRVLLIQTLLQLQDEIAD 674
DB 1267 E-----KAKGTLAEANAD-----LATELRSVNSSRQEND 1295
QY 675 -----YGLGMLKLSIPLTSLHSVELVPLCLRRSDVQSESG-----SPTDDNKDQA 721
DB 1296 RRRKQAESQIAELQVKLAIE-----RARSELOEKCTK-----LOQEAENITNQLEEASLKASA 1349
QY 722 APEDNEVDQEFLEKLETSFEFFLTSEEKLIQILTALCHRLTMYTSVQDHMETRQQSAELW 781
DB 1350 AVKASNNMESQL--TEAQQLLEETRQKLGLSKL-----RQIESE-- 1388
QY 782 KERLAVLKEEND-----KGAEKQK--RKEMEAKNKENGK----- 814
DB 1389 KEALQEOLEEDDEAKRNYERKLAETVTOQETIKKAEEDADLAKEEGKRLNKDIEAL 1448
QY 815 ---VENGLKTRDKRIIVKFEQVQVTEAEDMLSAVKSRRLLAIQAKREIEQEREMVKVL 871
DB 1449 ERQVKELTAQND---RLDKSKKIKOSELEDATIELEAQRTKYLELEK---QKNPDKILA 1502
QY 872 ERQAEERIRKHAAEKAFQGIKAKLVMRTPITGTR---NNRYWLFDEVEGLFI 928

DB 1503 BEKAISEIQAOERDTAEREAREKETKVLVSRELDDEAFDKIEDLENKRTKLQNELDDLAN 1562
QY 929 EKGWVHDSIDYRFNHHCCKDHTVSGDEDYCPSSKKANLGKNSMNTQHGTAATEVAVETTP 988
DB 1563 TOGTADKNV-----HELE-----KAKALESQALAEKQNELEL----- 1597
QY 989 KQGNLWFLCDSQKELDELLNCLHPQGIRESQLKER-----LEKRYQDIHSHLAR 1040
DB 1598 ----DLQLTEDAQLRLEVNMQALRSQFERDLLAKEEGAEKRRGLVKQLRQLETELDEER 1653
QY 1041 KPNLG-----LKSCDGNQELNLFRLSDLIIEVATRLQ---KGGIGVVEETSEFEA 1086
DB 1654 KORTAAVASKKLEGLDKIEIETTMEMHNKVEDALKHAKKLQAQVKDALRDAEEAKAAKE 1713
QY 1087 RVISLEKLDGREGCVIALQASVIKKFLOGFWAPKQREKLOSEDSAKTEE----- 1136
DB 1714 ELQALSKEAD-GK-VKALEAEVL-QLTEDLASSERRARAETERDELAEEIANNANKGSL 1770
QY 1137 -VDEEKNVVEEAKVASALEKWKTAIREAQTFRSMHVLGLMGLDACIKWDMSAENARCKVCP 1195
DB 1771 MIDEKRL--EARTATLEE---LEEQSNSEV-----LLDRAAARQLQIEQLTTELAN 1819
QY 1196 KKGEDDKILCDECNKAFHLFCLRPALYVEPDGWMQCPACQATARRNSRGRNYTESAS 1255
DB 1820 EKSNSQK---NENGR-----ALLERQNKELKAKLAEIETAQR-----TKVKAT 1859
QY 1256 EDSDEDSDESEEEEEEEEEEDYEVAGLRLRPKRTIRGKHSVIPPAARSGRPGKK-PH 1314
DB 1860 TATLEAKTARVEQLENEGKS-----RLLOQK-----ANKMDKKIKE 1897
QY 1315 STRRSQKAPPVDD--AEVDEL---VLQTKSSRRQSLELQK-----CEIL 1356
DB 1898 LTMWIEDERRHVDHQKEQMDKLNRIKLLKNLDETEELQKQTKRKYQRECEDMI 1955
RESULT 24
A41604
myosin heavy chain, smooth muscle, long splice form - rabbit
N:Contains: myosin ATPase (EC 3.6.4.1)
C:Species: Oryctolagus cuniculus (domestic rabbit)
C>Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 19-Apr-2002
C:Accession: A41604; A33501
R:Babi,j, P.; Kelly, C.; Periasamy, M.
Proc. Natl. Acad. Sci. U.S.A. 88, 10676-10680, 1991
A:Title: Characterization of a mammalian smooth muscle myosin heavy-chain gene: complete
A:Reference number: A41604; MUID:92073350; PMID:1961735
A:Accession: A41604
A:Molecule type: mRNA
A:Residues: 1-1972 <BAB>
A:Cross-references: GB:M77812
R:Nagai, R.; Kuro-o, M.; Babi,j, P.; Periasamy, M.
J. Biol. Chem. 264, 9734-9737, 1989
A:Title: Identification of two types of smooth muscle myosin heavy chain isoforms by cDN
A:Reference number: A33501; MUID:89255535; PMID:2722872
A:Accession: A33501
A:Molecule type: mRNA
A:Residues: 1455-1972 <NAG>
A:Cross-references: GB:J04833; NID:g165519; PIDN:AAA31407.1; PID:g165520
A:Experimental source: smooth muscle
A:Note: examination by Southern blotting for the regions of difference between this isofo
ative splicing
C:Superfamily: myosin heavy chain; myosin motor domain homology
C:Keywords: actin binding; alternative splicing; ATP; coiled coil; hydrolase; methylated
F:88-771/Domain: myosin motor domain homology <MMOT>
F:178-185/Region: nucleotide-binding motif A (P-loop)
F:559-572/Region: actin binding #status predicted
F:633-647/Region: actin binding #status predicted
F:844-1938/Domain: coiled coil #status predicted <COI>
F:844-1284/Region: S2
F:1285-1972/Region: light meromyosin
F:1939-1972/Domain: carboxyl-terminal <CBT>
F:129/Modified site: N6,N6,N6-trimethyllysine (Lys) #status predicted
F:184/Binding site: ATP (Lys) #status predicted

F;701,711/Active site: Cys #status predicted

Query Match	3.5%;	Score 279.5;	DB 1;	Length 1972;
Best Local Similarity	19.1%;	Pred. No. 7.1e-05;		
Matches	266;	Conservative 219;	Mismatch 503;	Indels 403; Gaps 60;

Qy	75	LLKEEPAWYEKL-VLEMVHHTASLEKLVDTAMLEMTK-----YAVGECDFEVGKEK	128
Db	801	LARAFAKROOQLTAMKVIQNRCAAYLKLNRQWWRLLFTKVKPILLOVTRQSEENQAKED	860
Qy	129	MLVKV-----IVKIHPLEKVDDEATEKKS-----DGACDPSDSDKENSQIAQDQHQK	176
Db	861	LQIKERQOAKAESELQELQOQKHQTQISEKNLQBOQAEETLYAEAEEMRVLAQKQEL	920
Qy	177	ETTVKDEGRRESINDRARSPRKPLTSLKKGERKWAPPKPLPHKYDVKLQNEDKIISNV	236
Db	921	EEILHEARLEEBEDRGQ-----LQAEK-----KMAQWMLDLEEQLB-----	961
Qy	237	PADSLIRTERPPNKEIVRYFIRHNALRAGTCENAPWVVEDELVKKYSIPSKFS---DFLL	293
Db	962	-----BEAARQKLQ---LEKVTAAEKIKKLEDDILVMDQNNKLSKERKLE	1005
Qy	294	DPYKYMTLNPSTKKNKTGSPDRKPSKSKTDNSLSPLNPKLWCHVHLKKSLSG---S	349
Db	1006	ERISDLTNNLAEBEKA-----KNLTKLKNKHESMISEL-----EVLKKKEKSROBLE	1054
Qy	350	PLKYNSKNKSPBEHLEEMKMMSPNKLHTNFHIPKGGPPAKP-----GHSDKPL	402
Db	1055	KLKFKMDGEASDLHEQIADLQAQIAELKMQL-----AKKEESLQAALARLEDETS	1104
Qy	403	KAKGRSGI-----LNGKSTGNSKSPK-----GLUKPTKTKQMTLMDMA	444
Db	1105	QKNALAKKIRELEGHISLOEDLDSEARAARKAEKQKRDLEEALKTELED---TLD-T	1161
Qy	445	KGTKMTRAPNNSGT-----PRTSS-----KPHKHLPPAALHLIAYVKENKOR	488
Db	1162	TATQOELRAKREQEVTVKALDETSHEAQVQEMQKQTQVVEETQEOLEQPRAKAN	1221
Qy	489	EDKRSALSCVISTARLLSSDDRARLPEELRSVLQKRYELLEHKHKRWASMSQOEKYLK	548
Db	1222	LD-----KTKQTLKEN-ADLAGELRVLQAKQE-VEHKK-----KLE	1258
Qy	549	KQREELKKLKEKAKERREKEMLEBSQKRYEDELTKGNLPAFRLVDPTEGLPNTLFG	608
Db	1259	VQLQELQSKGSD--GERARAEINDKVHKLQW-EVESVTG-----MLSEAEGKAIKLAK	1308
Qy	609	DVAMVVEFLSCYGLLPDQAQYPITAVSLMEALSADKGGFYLNRLVLQITLQTLQTLQ	668
Db	1309	EVASLGSQLOQTQELQOETRQKLVNVTKURQURDEBNS-----LOBQLD	1353
Qy	669	DETAEE-----DYGELGMKLSEIPLTLHSVSELVRLCLRSDSDVQSESGSDTD	715
Db	1354	EEMAEKQNLERHISTNLQISDSKKLQDPASTVESLEE-----GKRFQKEIESLTOQ	1407
Qy	716	DNKDSAAFE-----DNEVQDEF-----LEKLETSEFFELTSBEKLOILT	754
Db	1408	YEEKAAAYDKLEKTKNRLQOELDDLVDLQNRQLVSNLSK-KQKQPDQLLAEEK-----	1461
Qy	755	ALCHRIILMTYSVQDHMETROOMSABL-WKERLAV-----LKEENDKKRAEKQKRKE	804
Db	1462	-----NTSSKYA-----DERDRAEAAREKETKALSARALEEALEAKEELERTNMQLKAE	1512
Qy	805	MEAKNKENGKVENGLGKTRKRVKPEPOVD-----TEADMISAVKSRRL-----L	852
Db	1513	MEDLVSSKQDVGNVHELEKSKRAL---ETQMEEMKTOLEBEDELQATEDAKURLLEVNMQ	1570
Qy	853	AIQAKKEREIQER-----EMKVKLQRO-----ABEERIRKHAAKAEAFQEGIAKAKL	900
Db	1571	ALVKVQFERDLQARDEQNEEKRRQLQRLHEVETELEDERQALAAAKKLQGLDKOLE	1630
Qy	901	VMRTPTI-GYTRNNHNVWLSDEVPGFLFIKGNWVHDSIDYFRNHCHKDHTVSGDEDYC--	957
Db	1631	LOADSAIKGRE-----EAIKQLLKQAOAKMD-----FORELEDAASRDELIFATA	1675

```

QY 958 -PRSKAVLGNKSNVNTQHTAT-----EVAVETTTPKQGNLWFLCDSQKE 1003
    |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
DB 1676 KENEKKAASLEADLMQLQEDLAAARARQADLEKEELAEELASSLSGRNA--LODEKRR 1733
    |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 1004 LDELLNCLHPQIGRESQLKERLEKRYQDI IHSIHLARKPNLGLKSCDGNQELNLFRLSDL 1063
    :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
DB 1734 LE-----ARIAQLSEEEEB-----QGNMEAM-----SDR 1758
    :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 1064 IEVATRLKGGVGYVEETSEFEARVISLEKLKDFECVIALQASVIKKFLOGFMAPKQKR 1123
    |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
DB 1759 VRKATQ-----QAEQLSNELAT-----R 1777
    :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 1124 RKLQSDSAKTEEVDEEKVVEAKVASALEKWKTAIRE--AQTFSRMVLVGLMDACIK 1181
    |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
DB 1778 STAQKNESAR--QQLERQNK-----ELSKLQEMEGAVKSKFKSTIAALEAKIAQLSEQVE 1831
    :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 1182 WDMAENARCKVCPKGGDDKLILDCENKAFHLFCLRLPALYVEPVDGSMQCPACOPATAR 1241
    :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
DB 1832 QEAREKQAAKAL--KQDKK-----LKEMLLQVEDERKMAEQYKEQAEK 1874
    :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 1242 RNSGRNRYTE--ESASESDDEDESEEEEEEEEDYEVAGLRRLPRKT--IRGKHS 1297
    |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
DB 1875 GNKVKQLKQLEBAEESQRINARRKLQRELDEATESNMGREVNALKSLRRGNET 1934
    :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 1298 VIPAARSRR 1308
    |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
DB 1935 SFVTRRSRR 1945
    |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :

RESULT 25
S28261
centromere protein E - human
N:Alternate names: centromere 312K protein; kinesin-related protein CENP-E
C:Species: Homo sapiens (man)
C:Date: 03-Mar-1994 #sequence_revision 03-Mar-1994 #text_change 09-Jul-2004
C:Accession: S28261
R:Yen, T.J.; Li, G.; Schaar, B.T.; Szilak, I.; Cleveland, D.W.
Nature 359, 536-539, 1992
A:Title: CENP-E is a putative kinetochore motor that accumulates just before mitosis.
A:Reference number: S28261; MUID:93024922; PMID:1406971
A:Accession: S28261
A:Molecule type: mRNA
A:Residues: 1-2663 <YEN>
A:Cross-references: UNIPROT:Q02224; EMBL:Z15005; NID:G29864; PIDN:CAA78727.1; PID:G29864
C:Genetics:
A:Gene: GDB:CENPE
A:Cross-references: GDB:361164; OMIM:117143
A:Map position: 4q24-q25
C:Superfamily: centromere protein E; kinesin motor domain homology
C:Keywords: ATP; coiled coil; microtubule binding; mitosis; nucleotide binding; P-loop
F:7-335/Domain: kinesin motor domain homology <KMO7>
F:86-93/Region: nucleotide-binding motif A [P-loop]
F:486-2183/Domain: coiled coil #status predicted <COI>
F:92/Binding site: ATP (Lys) #status predicted

Query Match 3.5%; Score 279.5; DB 1; Length 2663;
Best Local Similarity 19.1%; Pred. No. 0.0001;
Matches 333; Conservative 307; Mismatches 608; Indels 491; Gaps 79;

QY 39 EARLERYSERIWTCKSGSOLTHKEAWEEBQEAELLKBEFPWAEYKLVLEMVHNTAS 98
    |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
DB 788 ESRVQGLEIEGKTKDDLATQSNYKTDQEFQNFKTLHMDPEQY-KWVLE---ENERM 843
    :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 99 LEKLVDPAWLEIMTKYAVGECDFEVEKEMKLVKI-VKTHPLSKVDDEATEKKS DGACD 157
    |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
DB 844 NOEIVN-----LSKEA--QKFQSSLG---ALKTELSYKTQELQEKTRVEQERLNE---- 888
    :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 158 SPSSDKENSQIAQDHQKKEVTVKVEDSGRESINDRARRSPRKLPT-SLKKGERKWAAPPK 216
    |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
DB 889 -----MBQLKEQLENRSDPSVREKTLITEKLQOTLEEVKTLUTQEKDLDLQJQES 940
    :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 217 FLPHKYDVKLQNEDKIISNPADSLIR-----TERPPNKETVYRIRNALRAG 265
    :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :

```

Db 941 LOIERDQLSDIHDVTNNMIDTQEOQLRNALESILKQHOETINTLKSKISEEVSRLHMEEN 1000
Qy 266 TGENAPVWVEDELVKYSILPSFSDFLDPDYKMTLNPSTKRTKNTGSPDRKPSKSKTDN 325
Db 1001 TGE-----TKDBFOQKMGVIDKKQD--LEAKNTQTLTADVKNEJIEQORKIFSLIOBKN 1053
Qy 326 ----SSLSPLNKLWCHVHLKSLSGSLKVKNSKNSKSPBEHLEEMKMSPNKLHTNF 382
Db 1054 ELQOWLESVIAEKEQKTDLKENIE---WTIENQBELRLGLLQKQOEIVAOEKRNA-- 1108
Qy 383 HIPKGPAPKPGKGSIDPLKAKGSKGILNCKSTGN-----SKSPKK-----GLKTPK 432
Db 1109 -IKKEGELSRTCORLAEEVEKLEKSKQLOEQOQLLVQBEEMSEMQKINEIENLKN-E 1166
Qy 433 TRKQKMTL-----LDMAK-----GTQKMTAPRPNGGTTPR----- 462
Db 1167 LKNKELTLHEMETERLELAQKLNENYEVKSYTKERKVLKELQKSFETERDHLRGYIREI 1226
Qy 463 --TSSKPHKHLPPAALHLIAYKENKDRED--KRSALSCVISKTLARLSSSED----RARL 514
Db 1227 EATGLOTKEELKIAHIHL-----KEHOETIDELRRS---VSEKTAQIINTQDLEKSHTKL 1278
Qy 515 PEELRSLVOKRYELLEHKRWASMSB-----EQRKEYLKKKR 551
Db 1279 QEEI-PVLHEQELLPNVKVSQETQTMNELLTQESTTKDSTTLARIEMERLRLNEKF 1337
Qy 552 BELKKLKEKAKERRE-----KEMLERLEKQRYEDQBELTKGNLPA 592
Db 1338 QESQBEIKSLTKRDNLAKTIKEALEVKHQDLKEHIRETLAKIQESQSKQEQSLNKE--- 1394
Qy 593 FLVDTPEGLPNTLPGDVAMVVEFLSCYSGILLPDAQYPIITAVSL-----MEALS 642
Db 1395 -----KDNETTKIVSEMOFPKPKOSALL-----RIEIMLGSLKQLQESHDEKNSVA 1441
Qy 643 ADKGFLYLRNVLLTQLLQTLQ-----DEIAEDYGELGMKLS 683
Db 1442 KEKDDLQRLQVLOESDQLENIKEIVAKHLETEBELKVAHCCLEKEQETINELRVNLS 1501
Qy 684 E-----IPLFLHVSVELRVLCLRRSDVQSESGSDTDNDKDSAAFEN----- 726
Db 1502 EKETEISTIQKLEAINDKL-----QNKIQEIEYEEQNLNKOISEVQNVNVELKQFKEH 1556
Qy 727 -EVQDEFLEKLTSEFFELT-----SEKLOJLTALCHRLMTYSVDHMETRQOWSAEL 780
Db 1557 RKAOSALQSIE-SKMLEUTNLQSSQSEIQIMIKEEM---KRVQEAQIOTERDOLKEN 1612
Qy 781 WKERLAVLKEENDK-----KRAEKQKRKEM---EAKNKENGKVENGLGKT 822
Db 1613 TXEIVAKMKESQEKYQFLKMTAVNETQEKMCIEIHLKEQFPETQKLNLENIETEN---- 1667
Qy 823 DRKGRIVKPEPOVDTAEADMI SAVKSR---RLLATQAKKERIOBREMVKVLEROAEBER 879
Db 1668 -----IRLTQILHENLEEMRSVTKERDRLRSVEETLKVVERDOLKENLRETITRDLEKQE 1721
Qy 880 IRKHAAAKAKAQEGITAKAK-LVMRT-----PIGTRDNRNRYLWLSDEVPGFLIEKGWVH 934
Db 1722 ELKIVHMLHKEHOETIDKLRGIVSEKTEINSNQKDLEHSNDALKAQD---LKIQE--- 1774
Qy 935 DSIDYRFNH-HCKDHTVSGDEBYCPRSKKANLGKASMMTQGTATAVETTTPKQGN 993
Db 1775 ---ELRIAHMLHKEQETIDKLRGIVSEKTKLSNQKDLNSNA----- 1816
Qy 994 LWFCLDSQKELDELNCLH-----PGITRESQJK-----ERLEKRYODIHSIHLARKNVL 1044
Db 1817 -----KLQEKIQELKANEHQILITLKDDVNETQKQVSEMEQLKQIKD--QSLTSLK--L 1866
Qy 1045 GLKSCDGNQELNFLRSLDIEVATRL-QKGGIGYVEETSEFARVISLEKLDQFGECVTA 1103
Db 1867 EIENLNLAQ-----LHENLEEMKSVNKRDNRLRRVEETLKLBR-----DQLKESQOETKA 1917
Qy 1104 LQASVKKPFLQGMAPKQKRR---KLQSDSDSAKTEBEVDEBKQWBEAKVASALEKWTAT 1160

Db 1918 RDLEIQELKTARMLSKHEKETVOKLREKISEKTIQIOTDKLDKSK-----DELOKKI 1972
Qy 1161 REATFTSRMHVLGMLDACIKWMSAENAR---CKVCPKKGEDDDKLIILDCBNKAFHLFC 1217
Db 1973 QELQK-KELQLLR-----VKEDVNMHSHKINEMEQKKQFEPNYLCKCEMDN-----FQ 2020
Qy 1218 LRPALYEPDCEWOCPCAPQATARNRGRNVTBESASEDSEDDSEDEEEEBEEEEEE 1277
Db 2021 LTKKLHESLE-EIRIVA-----KERDLRLRILKESLKMWR 2053
Qy 1278 DYEAGLRLPRKTIIRGKHVSIPPAARSGRPRGK-----PHSTRSQPPAPPVDDAEV 1331
Db 2054 DQFIATTLREMIARD-RQNHQV-----KPEKRLSDGQOHLMESLREK-----SRI 2098
Qy 1332 DELVLOTKRSS-----PROSLELOKCEIHLKI---VKYRFSWFRBPVITDEAED 1379
Db 2099 KELL---KRYSEMDDHYECLNRLSLDLEK-BIEFHRIMKKLKYVLSY-----VTIKBEEQ 2149
Qy 1380 YDVITVHPMDF-----QTVQNKSCSGSYR-----SVQEEFLTDMK 1413
Db 2150 HECINKFEMDFIDEVEKOKELLIKIQLHQDCDVPSERLDLKLNQNDMLHIEELKD--- 2207
Qy 1414 QVFTNAEYVNCRGSHVLSVMVTECLVYLLHKLHPGHPYVRRKRKKPPDRLAEDEGDSE 1473
Db 2208 --FSESEFPSS-----IKTEFQQVL-----SNRKEMTQPLEEWLNRFRDIE 2245
Qy 1474 PEAVQOSRDER-----RSREAEI-QEWLODTSLSYSAKINS 1509
Db 2246 KLKNGIQKENDRICQVNNFFNRIIMNTESTEFERSATISKEWEQDLKSLKEK-NEK 2303

RESULT 26
A1928
cag island protein - Helicobacter pylori (strain J99)
C:Species: Helicobacter pylori
A:Variety: strain J99
C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 09-Jul-2004
C:Accession: A71928
R:Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.;
Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.;
Nature 397, 176-180, 1999
A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric path
A:Reference number: A71800; MUID:99120557; PMID:9923682
A:Accession: A71928
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-1819 <ARN>
A:Cross-references: UNIPROT:Q9ZLV0; GB:AE001481; GB:AE001439; NID:g4155005; PIDN:AAD0604
A:Experimental source: strain J99
C:Genetics:
A:Gene: orf13/14

Query Match 3.5%; Score 279; DB 2; Length 1819;
Best Local Similarity 19.5%; Pred. No. 6.8e-05;
Matches 316; Conservative 229; Mismatches 54; Indels 530; Gaps 70;

Qy 139 PLEKVDKATE-----KKSDGACDCTSPSSDKNSQIAQDHQKKEIVKVEDGR 186
Db 149 PQDLSNEEATENAHFEDLLKESTESSDNLNDPTESSDNLNDPTETKTQTKTHFDEDK 208
Qy 187 RESINDRARRSPKLPSTSLKKGKRWAPPKLPHPKYDVKLQWEDKIIISNVPADSLIRTER 246
Db 209 PEEITDDSN-----DQEIIGSKK-----IIGGIIVAVLI-VII 243
Qy 247 PPNKEIVRYFIRHNALRAGTGENAPWVVEDELV-----KKYSLPSPKSFDFLL 293
Db 244 LFSRSIFHYFVPLEDKSRFSKDRNLYVNDQIRQRYNRLIKERNEKGNMIDKNLFFND 303
Qy 294 DP-----KYMTL-----NP-----STKKNYTGSPDRKP 317
Db 304 DPNRTLYNLNIAETEDKNPLRAFYECTISNGNVEECLKLIKDKKLQDQMKLTBAYNDC 363
Qy 318 SKSKSTDNSSLSSPLNPKLWC-----HVHLKKSLSGSLPKVKNS-----KNSKSPBEHLEE 368

QY 378 LHTNFIPIKGPAPKPGHSDKPLKA--KGRSKGLNGQKSTGNSKSPKGLKTPKTOM 435
Db 374 -----KEADEKKCEBAANKKEAAEKKKCEKAAKERKEAAEKKKCEAAKKKEKAA 425
QY 436 KQMTLLDMAKGTQKMTAPRNSGGTPTRTSSKPHKLPPAALHLIAIYYKENKDRED--KRS 493
Db 426 ERKKCEELAKNIKAAEKKKCEAAKKEAAERKKCEELAKKI KAAEKKKCEETAKKG 485
QY 494 ALSVSKTARLLSSEDRARLPPELRSVLQKRYELLEHK-----KRWASMSSEORKEYL 547
Db 486 KEVAERKKCEELAKKI KAEIKKKCKLAKKEKETAETAEKKCEKAAKKEAAEKKKCEKA 545
QY 548 KKRSEL--KKLKEKAKERRE-----KEMLERLEKQK-----RYE 581
Db 546 AKKRKEAAEKKCEKSAKKKEAAEKKKCEKAAKERKEAAEKKKCEAAKKEKVAERKK 605
QY 582 DOELTGKILPAPFLVDTPEGLPNTLFGDVAMVVEFLSCYSGLLLPDAQYPTITAVSLMEAL 641
Db 606 CELAKK-----IKKAAEKKKCKE-----NAKKEKAA 633
QY 642 SADKGGFLYNRVILLQTLQTLQDLQBIADYBGLG-----MKLSEIPLTLHVSSELV 696
Db 634 EREKCG-----ELAKKIKKAAEKKKCKLAKKEKETAEKKKE 671
QY 697 RLCLRRSDVOEBSGSDTDNDKDSAAFED--NEVQDFLEKLETSEFPFELTSEKQLIILT 754
Db 672 KAAKRRKEAAEKKKCEAAKKEAAEKKKCEAAKKEAAERKKCEELAKKI KAAAK 731
QY 755 ALCHRLIMTYSVDHMETR-----QOMSAELWKERLA----- 786
Db 732 KCKKLLAKKKKGERKKLKGNGKKGKALKKEKKCEELAKKAAEKKKCEAAKKEKAA 791
QY 787 -----VLKENDKGRABK--QKRKEM-----EAKNKENGKVENGL 819
Db 792 EKKCKEKTAKKKEAAEKKKCEKTAKKKEAAEKKKCEKAAKKEAAEKKKCEKTAKKR 851
QY 820 GXTDRKRVKPEQVDTAE--DMISAVKSRLLAIQ-----AKEREIOEREM---- 867
Db 852 KETAEEKKCEAAKRRKQAAEKKKCEKAAKKGKKEAAEKKKCEAAKKEKELAEKKCEBA 911
QY 868 ----KVKLRQAEERIRKHAAEKAFOEGIAKAKLVNRRPTPIGTDNRNHNRYLWLFSDV 923
Db 912 AKKEEVAERKKCEELAKKI KAAEK-----KCKKLLAKKCKKAGEKN----- 954
QY 924 PGLFTEKGWVHSDIDYRFNHHCQHTVSGDEDYCYPRSKANIGKNASMMNTQHTATEVA- 982
Db 955 -----KLKKGAGK-----KKCKKGLGKSKRAAEKKKCEAAK 988
QY 983 --VETTPKQGNLWFLCD----SQELDELNCLHPQGIRESOLKERLEKRYQDI IHSI 1036
Db 989 KEKEAATKK-----CBERAKKCEAAEKKKQCEB-----RAKGLKEAAEQ----- 1028
QY 1037 HILAKPNLGLSKDGNQOELLNLF--RSDLIETATRLQKGLGVVBETSFFEARVISLEKL 1094
Db 1029 -----KQCEERAKKLKEAAEKKKCEERAKKLKEAA-----EQKCEERAKKLKEA 1073
QY 1095 KDFGECVIAL-----QASVTKFLQGMAPKQKRKLQSDSASAKTEBEVDEKKNVBEA--- 1147
Db 1074 AEKKQCEERAKKEAAEKKKQCEERAKKLKEAAEKKKQCEERAKKEAAEKKKCEERAKR 1133
QY 1148 -----KVASALEKWKTAIREAQTFSRMHVLGMLDACIKWDSANARC----- 1191
Db 1134 EKEAAEKKKCEAAKKEKATEKQK-----CAEAAKKEKAAEKKKCEAAKKEK 1183
QY 1192 ----KVC-----PKGEDDKLILCECNK-----AFHLFLCLRPALYVDPDGEWQCPACOP 1237
Db 1184 EAAQKKCADLAKKEQPAEMKKCEAAKKEAAEKQKCAKAAKKEAAEKK--KCAE 1241
QY 1238 ATARNRSRGNVTESASDESDSDEEBEEBEEBEEDYEVLRLR--PRKTIROK 1295
Db 1242 AAKK-----EQEAAEKKKCEAAKKEKKEKKEKKEKKEKKEKKEKKEKKEKKEKKEK 1292

QY 1296 HSVIPP-----AARSRRPGKKPHSRRSQPKAPPVDDAEVDELVLQTKRSRRSQSLELOK 1351
Db 1293 EAAARLKKCAIIATAKAAKAAEKKCEKLA-KA-----KEAIEWKKCAKLAKKREAEKKK 1346
QY 1352 CEEILHK 1358
Db 1347 CAKLAKK 1353
RESULT 28
T21174
hypothetical protein F20G4.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T21174; T25113
R:McMurray, A.
submitted to the EMBL Data Library, June 1996
A:Reference number: Z19385
A:Accession: T21174
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-2020 <W1>
A:Cross-references: UNIPROT:Q19658; EMBL:Z75538; PIDN:CAA99841.1; GSPDB:GN00019; CESP:F20G4.3
A:Experimental source: clone F20G4
R:McMurray, A.
submitted to the EMBL Data Library, June 1996
A:Reference number: Z19982
A:Accession: T25113
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-2020 <W12>
A:Cross-references: EMBL:Z75550; PIDN:CAA99931.1; GSPDB:GN00019; CESP:F20G4.3
A:Experimental source: clone T22C1
C:Genetics:
A:Gene: CESP:F20G4.3
A:Map position: 1
A:Introns: 133/3; 155/2; 195/2; 270/3; 325/1; 1023/3; 1072/2; 1125/2; 1207/3; 1248/3; 1313/3
C:Superfamily: myosin heavy chain; myosin motor domain homology
F106-798/Domain: myosin motor domain homology <MMO>
Query Match 3.5%; Score 275.5; DB 2; Length 2020;
Best Local Similarity 19.5%; Pred. No. 0.0011;
Matches 298; Conservative 241; Mismatches 491; Indels 501; Gaps 72;
QY 11 PLVNP---LPGEPPFTIPHTQEAFTREEYEARELRSERLWTCSTGSSQLTHKEA-- 65
Db 760 PDVNPAGMFGDKNAVYRIVQYLEV-----DANLFR-----IGQSKIFFRSGVI 802
QY 66 --WEB--BOEVAELLKEFFPA-----WY-----EKLVLVMVHHNTASLEKLVDTAW 107
Db 803 AEFEENRQKLSALI--ESFQACRCGLGRRVVRREQEVAKILQRLGLAWRLREWQM 861
QY 108 LEIMTK----YAVGECDFEVGKERKMLKVKIVKI-----HPLEKVDDEATEKKS 152
Db 862 WRLLTKVKPLLEVTNKBELIAEREQELKVTAELKRSRVFISDYKQKQKMDERLVLKT 921
QY 153 DGACDPSDDK-----ENSSQIAQDHQKKTETVVVKEDEGRESINDRARSPRKLPTSLKKG 208
Db 922 --RLDAESRAEATFEERSMAARDELEGILEEVSKLEITEEQKAKK----- 967
QY 209 ERKWAPPKFLPHKYDVKLQNEDKIISNVPADSLIRTPPNKEIVRYFIRHNALRAGTGE 268
Db 968 -----ADUS-----ESRKLTEMVRLH-----EE 984
QY 269 NAPWVVEDELVKYSLPSKFSDFLLDPYKYMTLPNSTKRNKTSRDPKPSKSKTDNSSL 328
Db 985 N-----LEDE-----ERSRQKLLLEKNSI 1003
QY 329 SSPLPNKLWCHVHLKKSLSGSPLVKNSKNKSKSPPEH--LEEMKQMSNKLHNTNHIPIPK 387
Db 1004 ESRL-----KELEAQGLEDSGNKLSKEKKALEERCECDLSRLI----- 1043
QY 388 GPAPAKKPGKSHDKPLKAKGRSGIL-----NGQKSTGNSKSPKKGLKTPKTKMKQMT 439


```
QY 691 SVSELVRLCLRSVDQVESEGSDTDDNDKDSAAAFEDNEVDFLEKLETSFFELTSEBKL 750
Db 698 -----CLNKAFTDEERKCKLDPK-----LQSDILAKESLKAYKDCVSOAKT 701
QY 751 QILTALCHRLMTYSVQDHMETRQOWSAELWKRERLAVLKEENDKKRAEK-----OKRKEM 805
Db 702 EAEKKECEK-LUTPEAKKLEBEAKESVKAYLDCVSOAKTEAEKKECEKLLTPEAKKLE 760
QY 806 EAK-----NKGKQVENGIL-----GKTDR-KKRIYKF 831
Db 761 EAKSVKAYLDCVSRARNEKEKECEKLLTPEAKKLEQALDCLNKAFTDKERCKLKD 820
QY 832 EPQVTEADMI-SAVKSRLLAIAQAKEREIOEREMVKVLRQABEERIRKHAAAKA 890
Db 821 LPK-DLQKVLAKESVKAYLDCVSOAKTEAEKKECEKLLTPEAKLLEAKK-----SVKA 875
QY 891 FOEGIAKALVMRRTPIGTDNRHNYWLFSDVPGLFIEKGWVHDSIDYRFNHCKDHTV 950
Db 876 YLDCVSOAKTEAEKKECEK-----LUTPEAKLL----- 904
QY 951 SGDEYCPRSKCANLGNASMTNTOHGTAETAVETTPKQGQNLWFLCDSQKELDE---- 1006
Db 905 ---EEXAKESVKAYL-----DCVSOAKNEAEKKECEKLLT-ESKKLEAKKS 949
QY 1007 ----LNLCHLPQGISQLKERLEKRYQDIIHSIHLARKPNLGLKSCDG-----NQ 1053
Db 950 VRAYLDCV-----SOAKTEAEK-----KECEKLLTPEAKKLEQ 983
QY 1054 ELLNFRSLIEV-ATRLQKGLGYVEETSEFARVISLEKLDGFCVIALQASVKKF 1112
Db 984 QALDCLNKAFTADKKRCVK-----DLPKDQKVLAKESLUKAYDCVSKARNEKEKE 1037
QY 1113 LQGFMAPKOKRKLQSEDS-----AKTEE-----VDEKKMVEAEK----- 1148
Db 1038 CEKLLTPEAKKLEBAKSVKAYLDCVSOAKTEAEKKECEKLLTPEARKLLEAKESVKA 1097
QY 1149 ----VASAL-EKWTAIRAQFTSRMHVLLGLMDACIK-WDMSAENARC-KVCPKKGDD 1201
Db 1098 YKDCVSKARNEKEKECEKLLTPEAKKLEQOVLDCLNKAFTADKGRGVKDLPK--DLQ 1155
QY 1202 KLILDCENKAPHLCLRPALVEPDGEGQCPACOPATARRNSGRNVTES--ASEDSE 1259
Db 1156 KVLAKESVKAY-LDCVSRARNEKEKE-----CEKLLTPEAKKLEBAKESLUKYLKOL 1209
QY 1260 DDESEDEE-----EEEEEEEDYEAGLRPRKRTIRGKHSVIPPAA 1304
Db 1210 SQARNEERRACEKLLTPEARKLLEQEVKSIKAYLDCVSRARNEKEKECEKLLTPEAR 1269
QY 1305 -----SGRRPGKPHSTRSOPKAPPVDDAEVDELVLQTKR-----SSRRQ 1345
Db 1270 KFLAQVNLCLBKAGNEEERK--ACLKNLKP-----DLQENTLAKESLUKAYKDCLSOARN 1322
QY 1346 SLELOKCEIL---HKIVKRFSPFPBPVTRDEAEDYDVITHPMDFTQVONKSCGS 1401
Db 1323 EERRACEKLLTPEARKL-----EQVKSVKAYLDCVSR-----RNEKEKEC--- 1368
QY 1402 YRSVQEFLLDMKQVFTNAEYVNCRSHVLSCMVKTEQCLVLVLLKHLPCHGPYVRRKRKF 1461
Db 1369 ----EKLLTPEARKFLAKELQ-KOKAIKDCLKNAD-----PNDRAAIMKC 1409
QY 1462 PRLAED-----GSEPEAVQ-----SRDERSREAEIOEWLQD 1498
Db 1410 LDGLSDEELKYLQAREKAVADCLAMAKTDEBKRCQNLYSDLIQE 1456
```

RESULT 30

A59404

plectin [imported] - human

C:Species: Homo sapiens (man)

C:Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 09-Jul-2004

C:Accession: C59404; A59404

R.Liu, C.G.; Maercker, C.; Castanon, M.J.; Hauptmann, R.; Wiche, G.

Proc. Natl. Acad. Sci. U.S.A. 93, 4278-4283, 1996
A:Title: Human plectin: organization of the gene, sequence analysis, and chromosome loca
A:Reference number: C59404; MUID:96210632; PMID:8633055

A:Accession: C59404

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-4684 <STO>

A:Cross-references: UNIPROT:Q15149; GB:CAA91196; NID:gl296662; PIDN:CAA91196.1

C:Superfamily: plectin; alpha-actinin actin-binding domain homology; ribosomal protein s

Query Match 3.4%; Score 274; DB 2; Length 4684;

Best Local Similarity 17.8%; Pred. No. 0.00034;

Matches 277; Conservative 241; Mismatches 562; Indels 478; Gaps 56;

QY 25 IPHTQEAFTREEYV--ARLERYSERIWTCKSTGSSQLTHKE-AWEEQEVAEI----- 75

Db 1471 IKFISETLRRMEERLAEQQAER-----ERLAEVEAALKEQRLAEHAQA 1521

QY 76 -LKEEFPWYKLVLEWVHNTAS-----LEKLVDTAWLIMTKYAVGEBCD 121

Db 1522 QAEREAKELQRIQEEVVRREEAAVDAQQKRSIQEELQQLRQSEASESIQAKARQAEAE 1581

QY 122 ---FEVGEKMLVKVIKIHPLKVDDEATEKKSQACDPSDDKENSQIAQDHQKKET 178

Db 1582 RSLRIEE---IRVVEL-----QLEATERQGA----- 1608

QY 179 VKDEGRRESINDRARRSPKLTPTSLKKGKRWAPPKPLPHKYDVKLQNEDKIISNVA 238

Db 1609 ----EGSLQALRAAEAEAKQQAQEAERL-----RRQVDESORKQAEVEL 1654

QY 239 DSLINTERPPNKEIVRYFIRHNALRAGTGENAPVVEDELVKYSLPSKFSDFLLDPKY 298

Db 1655 ASRVKAEAEAREKQALQALELRLQAEAEERMLCQAEVERARQV-----Q 1701

QY 299 MTLNPSTKRKNTGSPDRPKSKKTDNSSLSPNKLWCHVHLKKSLSGPLKV----- 353

Db 1702 VALETAQSAEAEALQSKASFAEKT-----AQLESLQEETHVAQA 1745

QY 354 ---KNSKNKSPBEHLEEMKQMSPNKLHTNPHIPKGGPPAKPKGKSDKPLKAGRSKG 410

Db 1746 EAERRAQQAERAREAREAEERQLERWQLKAN-----EALRLQAEVLEVLQKSLQA 1797

QY 411 ILNGKSTGNSKSPKGLKTPKTKOMKOTLLD-----MAKGTQKMTPRNNSGGT 462

Db 1798 EAKQKEAEAREARRG--KABEQVRQRELAEOLEKORQAEQQAORLA----- 1847

QY 463 TSSPHKHLPPAALHLIAIYYKENKDRDKRSALSCVISKTARLLSSEDRARLPEELSLV 522

Db 1848 -----AEQELIRLAEQEGEQORQL-----EELARLQREAAAT 1884

QY 523 QKRYEL-LEHKRWASMSSEQRKEYL-----KKRBEELKKLKEKAKERBEKE----- 569

Db 1885 QKQLEAEALAKVRAEM-----EVLLASKAKAEESRSTSEKSKQRLAEAGREFELAE 1938

QY 570 -----MLERLEKQRYEDQELTGKINLPAPFLVDTPPEGLPNTLPGDVAMVVEFLSCYS 622

Db 1939 EAARLRLAEAEAKQRLAEEDAAQRAEAEV----- 1971

QY 623 LLLPDAQYPIAVSLMEALSADKGGFLYNRLVILLQTLQTLQDLQADYDGLGMLK 682

Db 1972 -----LAEKLAIGEATRLKTEAEIALKEKEAENERLRLAEDEAFQRRRL 2017

QY 683 SEIPLTHSVSELVRLCLRRSDVOBE-----SEGSDTDNKSAAAFED-----NEVDFEL 733

Db 2018 EE-----QAAQHKAIDIEERLAQLRAKASDELERQKGLVEDTLRQRQVEEIL 2065

QY 734 EKLETSEFFELTSEBKLQILTALCHRLMTYSVQDHMETRQOWSAELWKRERLAVLKEEND 793

Db 2066 -ALKAS---FEKAAAGKASLELELGR---IRSNAEOTLRSKEQAELEAARQALAEERR 2119

QY 794 KKRAE-----KQKKMEAKNKENGKVLNGKTKDKRKIRVFPQVDTE- 838

Db 2120 RREAERVQKSLAAEEEAARQKAALEEVERLKAKEVEARSRLRERAEQESARQLQAQEA 2179

Db 860 DXL-----SKVSDLSKNSKSEFEKENQKGAAILDLEKT----- 893

QY 991 GONLWFLCDSQKELDELNCLHPQIRE-SQLKERLEKRYQDIHSIHARKPNLGLKSC 1049

Db 894 -----C---KELKHQLQVQMENTLKEQKELKKSLEKE-KEASHQLKLE----- 932

QY 1050 DGNQELLNPLRDLIEVATRLQ-----KGGLYGYEETSEFEARVISLEKLKDPGEC 1100

Db 933 -----LNSMQQLIOAONTLKONEKEBOQLQGNINELKQSEQKKK--QIEALQ--GEL 982

QY 1101 VIA-LQASVIKKFL-----QGFMAPKQKRKLQS--EDSAKT----- 1134

Db 983 KIAVLQKTELENKLOQLTOAAQELAAEKEKISVLQNNYKESQETFKQLQSDFYGRESL 1042

QY 1135 -----EEVDEEKQVVEAKVASALEKWKTKATAREATFFRM 1169

Db 1043 LATRODLKSVEKLSLAQEDLISNRQIGNQNKLIQELTKATLEQDSAKKEQQLQERC 1102

QY 1170 HVLLGMLDACIKWMSAENARCKV-----CPKGGEDKLILCDECNKAPHLCLRPAL 1222

Db 1103 KALQDIQKEKSLKELVNEKSLKLAIEIKRCQKEITKL---NEELKSHKLESIK-EI 1158

QY 1223 YVDPG-----EWQCPA-CQPATARNRGRNYYTEESASEDSEDESEDESEEE 1272

Db 1159 TNLKAKQLLLOKLELQKADSLKAAVEQKRNQIILKQVKKSEELKKEFIKEAKL 1218

QY 1273 EEBEEDYEVAGURLRPRKTIRGKHSVIPPAASSRRPGKPKHSTRSQPPVDDAB-- 1330

Db 1219 HSEIKEKEV-GMKHEENEAKLTQITALNENLGTVKKEWQSSQRRVSELEKQTDLRGE 1277

QY 1331 --VDELVLQTKSSRQSLQKQ-----EELHLKIVKYR----- 1363

Db 1278 IAVLEATVQNNQDERRALLE--RCLKGEGIEKQTKVLELQKLDNTTAVQELGRENQS 1335

QY 1364 FSWPREPVTREAEYDVITHPMDFTQVQNKCGSGSVRSV 1405

Db 1336 LQIKHTQALNRKWAED-----NEVQNCMAGKGFVS 1366

RESULT 32

S37431

N:Alternate names: ankyrin B, 440K splice form; ankyrin-B; brain ankyrin; non-erythroid

N:Contains: ankyrin 2, short form

C:Species: Homo sapiens (man)

C>Date: 06-Jan-1995 #sequence revision 06-Jan-1995 #text change 09-Jul-2004

C:Accession: S37431; E39643; E39643; A40334; A49462; S14533; S14569

R:Chan, W.

submitted to the EMBL Data Library, September 1993

A:Reference number: S37431

A:Accession: S37431

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-3924 <CHA>

A:Cross-references: UNIPROT:Q01484; EMBL:Z26634; NID:G406287; PIDN:CAA81387.1; PID:G4062

R:Otto, E.; Kunimoto, M.; McLaughlin, T.; Bennett, V.

J. Cell Biol. 114, 241-253, 1991

A:Title: Isolation and characterization of cDNAs encoding human brain ankyrins reveal a

A:Reference number: A39643; MUID:91302466; PMID:1830053

A:Accession: A39643

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-2077 <OTI>

A:Cross-references: GB:X56957

A:Accession: B39643

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-1443,3585-3924 <OTT>

A:Cross-references: EMBL:X56958

R:Tse, W.T.; Menninger, J.C.; Yang-Peng, T.L.; Francke, U.; Sahr, K.E.; Lux, S.E.; Ward,

Genomics 10, 858-866, 1991

A:Title: Isolation and chromosomal localization of a novel nonerythroid ankyrin gene.

A:Reference number: A40334; MUID:92009921; PMID:1833308

A:Accession: A40334

A:Molecule type: DNA

A:Residues: 463-474, 'PE', 477-495 <TSE>

A:Cross-references: GB:M37123; NID:G178647; PIDN:AAA62828.1; PID:G178648

R:Chan, W.; Kordeli, E.; Bennett, V.

J. Cell Biol. 123, 1463-1473, 1993

A:Title: 440-kD ankyrinB: structure of the major developmentally regulated domain and se

A:Reference number: A49462; MUID:94075409; PMID:8253844

A:Accession: A49462

A>Status: preliminary; nucleic acid sequence not shown

A:Molecule type: mRNA

A:Residues: 1-3924 <RES>

A:Cross-references: EMBL:Z26634; NID:G406287; PIDN:CAA81387.1; PID:G406288

C:Genetics:

A:Gene: GDB:ANK2

A:Cross-references: GDB:127607; OMIM:106410

A:Map position: 4q25-4q27

C:Superfamily: ankyrin; ankyrin repeat homology

C:Keywords: alternative splicing

F:2-3924/Product: ankyrin 2, long form #status predicted <MAT>

F:2-1443,3585-3924/Product: ankyrin 2, short form #status predicted <MA2>

F:63-95/Domain: ankyrin repeat homology <AN01>

F:96-128/Domain: ankyrin repeat homology <AN02>

F:129-161/Domain: ankyrin repeat homology <AN03>

F:162-190/Domain: ankyrin repeat homology <AN05>

F:191-223/Domain: ankyrin repeat homology <AN06>

F:232-264/Domain: ankyrin repeat homology <AN07>

F:265-297/Domain: ankyrin repeat homology <AN08>

F:298-330/Domain: ankyrin repeat homology <AN09>

F:331-363/Domain: ankyrin repeat homology <AN10>

F:364-396/Domain: ankyrin repeat homology <AN11>

F:397-429/Domain: ankyrin repeat homology <AN12>

F:430-462/Domain: ankyrin repeat homology <AN13>

F:463-495/Domain: ankyrin repeat homology <AN14>

F:496-528/Domain: ankyrin repeat homology <AN15>

F:529-561/Domain: ankyrin repeat homology <AN16>

F:562-594/Domain: ankyrin repeat homology <AN17>

F:595-627/Domain: ankyrin repeat homology <AN18>

F:628-660/Domain: ankyrin repeat homology <AN19>

F:661-693/Domain: ankyrin repeat homology <AN20>

F:694-726/Domain: ankyrin repeat homology <AN21>

F:727-759/Domain: ankyrin repeat homology <AN22>

F:760-792/Domain: ankyrin repeat homology <AN23>

F:793-825/Domain: ankyrin repeat homology <AN23>

Query Match 3.4%; Score 272.5; DB 2; Length 3924;

Best Local Similarity 19.5%; Pred. NO. 0.00032;

Matches 348; Conservative 243; Mismatches 673; Indels 523; Gaps 81;

QY 19 EEPFTPIPTQAPRTREYEARELYSE--RIWTC-KSTGSSQLTHKEAWEEBEEVQEL 75

Db 1510 EEP-----GEPFEIVERVKEDLEKVNELRSGTCTRDESSVSSRSEGLVEEWIV 1562

QY 76 LKEFPAPVYKLVLEVMVHNTASLEKLVDTAWLEIMTKYAVGECDFEVGKMKMLKIV 135

Db 1563 SDEIEEARQAPLE-----ITEY-----PC-VEVRIDKEIKGKVE 1597

QY 136 K-----IHPLEKVPDEATEKSDGACDS-----PSSDKE-----NSSQIA 170

Db 1598 KDSGLVNYLTDLNTCVLPKEQLQTVQDKAGKCEALVGRSSEKEGKDIPTDETQST 1657

QY 171 QDHQKQETVVKEDGRRESINDRARRSPKPLTSLKGERKWAQPKFLPHKYDVLQNEED 230

Db 1658 QKHQKPSLGIKKPV--RRKLEKQKQKEEGQASAKAELK-----KGSSEE 1702

QY 231 KI-----ISNVPADSLIRTERPPN--KEIVRYFIRHNALRAGTGENAPWVVEDEL 278

Db 1703 SLGEDPGLAPFLPTVKATSLPIETPTGSIKDKVKALQKR-----VEDEQ 1748

QY 279 VKKYSLSKSFDFLLDPYK-----YMTLNPSTK--RKNTGSPDRKPSKKSKTDSNLSLS- 330

Db 1749 KGRSKLPVRVKGEDVPKTKTTHRPHPAASPSLKSERHAPGSPSPKTERHSTLSSAKTER 1808

QY 331 --PLNPKLVCHVHLKKSLSGSPKLVKXNSKNKSPPEHLEBMMQMGSPNKLHTNFHIP-----385
Db 1809 HPPVSPSSKTEKHSVPSPSAKTERHSPASSSSSTEKH-----SPVSPS-TKTERHSPVSS 1862
QY 386 ---KGGPPAKKPGKSHDK--PLKAKGRSKGILNGQKSTGNSKSPKGLKTPKTKMKQMTL 440
Db 1863 TKTERHPPVSPSK-TDKRPPVSPSGRTE-----KHPP-----1894
QY 441 LDMAGTKQKMTAPRNSGGTPR-----TSSKPHKHLPPAALHLIAYYKKNKREDKRSAL 495
Db 1895 VSPGRTEKRLPVSP--SGRTDKHPVSTAGTKEKHLVPSP-----SGKTEKQPPV 1942
QY 496 SCVISKTARLLSSEDRARLPEELRSLVQKRYELLEHKRWASMSERQKYLEKKREELK 555
Db 1943 S-PTSSTERI-----EETMSV-----REL--KAFQSGQDPSKHKHTGLFEHKSQK 1985
QY 556 KKLKEKAKERREKE---MLERLEKQKRYEDQELTGKNLPAFLVDPTEGLPNTLFG-----608
Db 1986 KQOEKGVKRVKEKGPILTORAQKTENQTIKRGQRLPVTGTAESKRGVRVSSIGVKKE 2045
QY 609 DVAMVVEFLSCYSGLLLPDAQYPTITAVSLMEALSADKGGFLYLNRVLVILLQTLQLLQ 668
Db 2046 DAAGGKEVLSHK--IPE---PVQSPPEESHRESE-----VPK 2079
QY 669 DETAEDYGEKGMKLSIPIPLTHSVSELVRLCLRRSD-----VOESESGETD-----DN 717
Db 2080 EKMADEQGMDLQISPDRTKSTDPSEVIKQELDNDKYQOFRLSSETEKAQLHLQDVLTS 2139
QY 718 KOSAAFEDNEVDFLEKLETSEFFELTSEEKLO---ILTALCHRLIMTYSVQDHMETRQ 774.
Db 2140 PFNTTFFLDYMKDEFPLALSGALDGSSELKNEGAVAGSPGSL-----MEGTP 2190
QY 775 QMSAE--LWKERLAVLKEENDKKRA-EKQKREMEAKNKENGKVENGLGDKTKRKRIKVF 831
Db 2191 QISSESYKHGELAETPEISPELSFSPKSEBQGTGETKESYKTE-----TTTEIRSEKE 2245
QY 832 EPOVDTEADMTSAVKSRLILAIQAKEREIOQREMKVKLERQAEERIRKHKAAAEKAF 891
Db 2246 HP-----TTKDIITGSGSERGATVTE-----DSETSTESQKEATLGSPKDTSPKRQ 2291
QY 892 QEGIAKALVM-RRTPIGTDRNHNRYLPSDEVPGLPIEKGWVHDSIDYRFNHCHCKDHY 950
Db 2292 DDTGSCSVALAKETPTGTEE-----AACDEGQRTFGSS--AHKT---QTDSEAQESTFA 2341
QY 951 SGDEDYCPRSKXKANLGNAMNTQHGTEAVAVETTPKQGNLWFLCDSQKEL-----DE 1006
Db 2342 TSDE-----TKALPIPEASVKTDGTESKPGQVIRSP-QGLEALPSPRSDSEVLSAVD 2394
QY 1007 LNCLEH-----PGIRESQLEKLEKRYQDIITHSIHLARKPNLG-----1045
Db 2395 SLAVSHKDSLEASPVLEDNSSHKTPDLSPLK---ESPCRDLSLESSPVPKMKAGIFP 2451
QY 1046 ----LKSCDGNQELLNFLRDLIEVATRIKQGLGVVETSPFEARVISLEKLDKPGECV 1101
Db 2452 SHFPLPAVAKTELL-----TEVASVRSLRLRDPDGSAAED-----SLE-----2490
QY 1102 IALQASVIKFKLQGFMAP-KQKRKLQSDSDAKT-----EEVDEEK-----1141
Db 2491 ----QTSLMSSGKSPSPDTPSPSESVSYVPTKTTDVTVPKPAVIEHCEABDDSENGEKK 2547
QY 1142 -----KWBEAKVASALEK-----WKTATREAOQTFRRMHVLLGMLDACIKWDM 1185
Db 2548 RFTPEBEMFKWTKIKWFDELEQAKQKDYKKEPKQESSSS-----SS 2591
QY 1186 AENARCKVC---PK---KEDDK---LILCDECNKAFHLFCLRPALEYVDPG-----1228
Db 2592 DPDADCSVDVDEPKHTGSGEDESVPVLTSTESRKYSSSESEPELAQLKKGADSGLLPE 2651
QY 1229 ---EMQCPACQAPATARNR-----GRNYTEESASEDSEDESEEEEEEEEEE 1275
Db 2652 PVIRVQPPFLPSSMDNSPPEVQPPVSVKQYTFK-MNEDTQEPGKSEBKEKSESIL 2710
QY 1276 EEDYEYAGLRLPRKTIIRGKHSVIPPAARSGR-----RPGKPHSTRRSQPKAPPVD- 1327

Db 2711 AEDRHAVSTEAEDRSYDKLNRDTPDKICDGHGCEAMSPSSSARPVSSGLQSGPTGDVDE 2770
QY 1328 -----DAEVDLVLQTKRS-----1341
Db 2771 QPVIYKESIALQWTEHKDTEGELDVSRAESQADCPSESFSSSSLPHCLVSEGKELDE 2830
QY 1342 -----SRROSLBLOKCEEILHKIVK-----YRFSW--PFREPVTRDEAEDYY 1381
Db 2831 DISATSSIQKTEVTKTDETFENLPKDCPSQDSSITQTDRFSMDVPVSDLAENDEL--YD 2888
QY 1382 DVITIPMDPQTQVONKCSGYSRSVQEFLLTDMKQVFTNAEVNCRGSHVLSCHWVKTSQCLV 1441
Db 2889 POITSP--YENVPQSOFSSSEESKTKTDANHTTSPHSSESVSYTITSPPVEDVVVASSSSG 2946
QY 1442 VLLHKLPHLPVYRRKRKXFPDLAEDEGS-----EP-----EAVGQ-----SRD 1482
Db 2947 TVLSKESNPEGQDIKNESQLESTLWEMQSDSVSSSEFPTMSATTTVVGQISKVIITKTD 3006
QY 1483 EDRRSREABIQEWLQDTSLYSAKINSKDHNCFMMLVNTQFCMALTDT 1529
Db 3007 VDSDS-WSEIRE---DDEAFEARVKEEOKIFGLMVDROSQGTTPDT 3049

RESULT 33
A45973
trichohyalin - human
C:Species: Homo sapiens (man)
C>Date: 03-May-1994 #sequence_revision 01-Mar-1996 #text_change 09-Jul-2004
C:Accession: A45973
R:Lee, S.C.; Kim, I.G.; Marekov, L.N.; O'Keefe, E.J.; Parry, D.A.D.; Steinert, P.M.
J. Biol. Chem. 268, 12164-12176, 1993
A:Title: The structure of human trichohyalin. Potential multiple roles as a functional
ed (cross-linking) protein.
A:Reference number: A45973; MUID:93280194; PMID:7685034
A:Accession: A45973
A:Molecule type: DNA
A:Residues: 1-1898 <LEE>
A:Cross-references: UNIPROT:Q07283; GB:I09190; NID:G292835; PIDN:AAA65582.1; PID:G29283
A:Note: authors translated the codon AGG for residue 1714 as Pro
C:Comment: Trichohyalin is a protein of the medulla of the hair and of the inner root sheath.
Covalent modifications to this protein include conversion of arginine to citrulline and
C:Genetics:
A:Gene: GDB:THH
A:Cross-references: GDB:136223; OMIM:190370
A:Map position: 1q21-1q21
C:Superfamily: trichohyalin; calmodulin repeat homology
C:Keywords: calcium binding; citrulline; EF hand; hair; tandem repeat
F:49-81/Domain: calmodulin repeat homology <EF2>

Query Match 3.4%; Score 272; DB 1; Length 1898;
Best Local Similarity 18.9%; Pred. No. 0.00015;
Matches 203; Conservative 188; Mismatches 417; Indels 264; Gaps 37;

QY 483 KENKREKRSALSCVISKTARLLSSEDRARLPEELRSLVQKRYELLEHKRWASMSSEEQ 542
Db 288 REREEEQQQQLR--REQQLRRKQEEERREERREERREERREERREERREERREERRE 345
QY 543 RKE-VLKKEKREBKKK---LKEKAKERREKEML-----ERLEKQKRYEDELTKQMLPAP 593
Db 346 RREQQLRREERREERREERREERREERREERREERREERREERREERREERREERRE 395
QY 594 RLVDTPGELPNTLFGDVAMVVEFLSCYSGLLLPDAQYPTITAVSLMEALSADKGGFLYLR 653
Db 396 -----LRR 398
QY 654 VLIVILLQTL-----QTLQDETAEDYGEKGMKLSIPIPLTHSVSELVR-LCLRRSDVQERS 709
Db 399 EQQLRREERREERREERREERREERREERREERREERREERREERREERREERREER 453
QY 710 EGSDDDDKDSAAFEDNEVDFLEKLETSEFFELTSEEKLOILTALCHRLIMTYSVQDH 769
Db 454 EQKHEQERREQLKREQBERDLKREETERHE--QERRKQOLK-----RDQ 499

Qy	770	METROQMSAELMWERLAVLKEENDKKRAEKOKRKEAKNGKENGKVENGLKTKDRKKRI-	828
Db	500	EEER-----RERWLKLEEB---ERRQERREQLRRQERREORLKRQBEERLQ	548
Qy	829	--VKPEPQVDTAEADWISAVKSRRLAIQAKKERIQEREMKVKLERQAEERIRKHAA	886
Db	549	QRLRSEQLRRQERL-----EQLLKREERLEQRERQRLKREQBERRQDLKRE--	601
Qy	887	AKAFQEGIAKAKLVNRTPITGDRNNHRYWLFSDVEVPLFTEKGWVHDSIDYRNFHCK	946
Db	602	-ERRQQLRKREQ-----EERLEQR--LKREVERLEQER-----R	635
Qy	947	DHTVSGDEYCYPRSKKAN--LGNASMTQHGTATEVAVETTPKQGNLWFLCDSQEL	1004
Db	636	DELRKREE---PEERRHLLKSEQEERHQLRREOQERORLKR-----EEREL	687
Qy	1005	DELLNCLHPQIGRESOL-----KERLEKRY-----QDIHSHLARKPNLGL	1046
Db	688	EQLKREHEERREQLABEEQOARERIKSRIPKWQMLESEADARQSKVLEAPOAGR	747
Qy	1047	KSCDGNQLLNLFLRSDL-----LEVATRLQKGLGVYEETSEFEA-----RVISLEK	1093
Db	748	AEAPQOEKRRRESELOWQBEERAHROQOEBEQRDRFTWQABEKSERGRQLSARPP	807
Qy	1094	LKDFGECVIALQASVTKFLQFMAPQKR--RLQSEDSAKTEVDREKMWVEAKVAS	1151
Db	808	LRSQRE--ROLRAEERQOERQFLPEEBEKEQGRORREKELOFLBEEEOQLORRRAQ	865
Qy	1152	ALEKWKTAIREAQTSRMHVLLGMLDACIKWMSAENARCKVCPKKGEDDKLILCDECNK	1211
Db	866	QLOEEDGLOEQERRQOE---QRDQKRWQLEBEERKRR-----	903
Qy	1212	APHLFCLRPALYE-----VPDGEWQCPACOPATARNRSGRNYTESASESDEDES	1263
Db	904	--HTLYAKPALQBLRKEOQLQOEBEELQREERKRRRQOERQYREEEOQLQOEBEQLL	961
Qy	1264	DEEBEEEEEEEDYEVAGLRLRPKTRIRGHVSIPPAARSGRRPKPHSTRRSOPKA	1323
Db	962	REERKRRQERQY-----RKDKLQOKEEQ-----LGEPEKRRQERREKRYRE	1009
Qy	1324	PPVDDAEVDLVLQTKRSRRRSLELQ--KCEIHLHKVVKYRFPFPVTRDEADYY	1381
Db	1010	ESELQOEBEQLREERKRRRQEWERYKQDELQOE-----EEQLREERK--	1057
Qy	1382	DVITHPMDFTQVKNKSCGYSYRSQVFLDMQVFTNAEYVNCRGSHVLSCMVTEQCLV	1441
Db	1058	-----RRLQERER--QYREEELOQEBEQLL--GEERETRRRQELERQYRKEBELQ	1104
Qy	1442	VLLKHLPGHPYVRKPKFPDLRADEGDSPEAVGQSRDEDRSRRAEIQ	1493
Db	1105	QBEELQLREPEKRRQERROCRBEELQOEBEQLLREERKRRRQELERQ	1156
RESULT 34			
B85431			
trichohyalin like protein [imported] - Arabidopsis thaliana			
C:Species: Arabidopsis thaliana (mouse-ear cress)			
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 09-Jul-2004			
C:Accession: B85431			
R:anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Spring			
Nature 402, 769-777, 1999			
A:Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.			
A:Reference number: A85001; MUID:20083488; PMID:10617198			
A:Accession: B85431			
A>Status: preliminary			
A:Molecule type: DNA			
A:Residues: 1-1432 <STO>			
A:Cross-references: UNIPROT:O23230; GB:NC_001268; NID:g7270600; PIDN:CAB80318.1; GSPDB:C			
C:Genetics:			
A:Gene: AT4g36520			
A:Map position: 4			

Query Match	3.3%;	Score 265.5;	DB 2;	Length 1432;
Best Local Similarity	19.9%;	Pred. No. 0.00021;		
Matches 284;	Conservative 212;	Mismatches 531;	Indels 399;	Gaps 65;
Qy	61	THKEAWEEBQEAELKEEPFAPWYEKLVLEMVHHTASLEKLVDTAWLMTKYAVGEEC	120	
Db	268	THPTGIPPPSRAAPLTKSDFRS-----SASNKTTGSGQSDVSS--SPTFFDVEVDA	318	
Qy	121	DPEVGKEMLKVKVIKIHPILEKVDDEATEKKSDGA--CDSPSSDKENSSQIADHQHKKET	178	
Db	319	NSAAVREANLXAE-AKL---KSAKELLERKRDVARTTSTPSSTK---NRKEEGKSSYA	370	
Qy	179	-----VVKED-EGRESIND-RARRSPRLPTSLKKGERKWAPPKFLPHKVDVQLQNEDKI	232	
Db	371	AGLPVVKYINIDVARKSLADKRGSKLSQGSATSDGNDW-----KEANNQYV	418	
Qy	233	ISNVPADSLITERPPNKE-----IVRVFIRHNALRACTGTGENAPWV-----	274	
Db	419	-----ELVRTLEPRNADENSGKDVSIPLNABEFFQOELTWAG--NVWDEKQRRRAKG	468	
Qy	275	--EDELVKKYSLPKSFDFLLDPYKMTLPNSTRKNTGSPDRKPSKSKTDSNLSL--S	330	
Db	469	DREDHEARK--LP-----KHSGT--RKLSSRHKHENKLAEKA	502	
Qy	331	PLNPKL--WCHVHLKKSLSGSPLKVKNSKNSKSPSEHLEEMKMKMSPNKLUHTPHIPKKG	388	
Db	503	PEEPKIEKSRHVEMGNLDPDHGGIVGH-RNLLKPEE-----NKLFT-----	542	
Qy	389	PPAKPGKHSKPLKAKGRSGILNGQKSTGNSKSPKGLKTPKTKMKQMTLLDMAKGTQ	448	
Db	543	---EKPAKQK-KELCEEEKTKRIQNQQO---DKKTHQKAAET-----NQBSCVVDWEONAR	590	
Qy	449	KMTRAPRSGGTPRTSPKPKHPLPAALHLIAYYKENKDKREDKRSALSCVISKTAARLSS	508	
Db	591	KUREALGNE-----STLEVSVELNGNGKMWKMSQSETKLNLEPKRMEE	634	
Qy	509	EDR---ARLPEELSLRVQRYELLEHHKRW-ASMSERQKEYLKKGRKEELK-----KLK	559	
Db	635	ETRIKEARLREENDRRERVAVEKAENEKRLKAALQEOKERKIKEAREKAENERRAVEAR	694	
Qy	560	EKA-KERREKEMLE-RLEKQKRYEDQELTGKNLPAFLVDTPEGLPNTLFGDVAMVWVEFL	617	
Db	695	EKAQERKMKQEQLLEQLKEAFKEBEENRRMREAFALQOEKER-----	738	
Qy	618	SCVSGLLLPDAQYPITAVSLMEALSADKG-GFLVNLRVLVILLQTLTLQTLQDEIAEDYG	676	
Db	739	-----RIKEAREKENERRRIKEAREKAELEQRLEKATLEQEKERQIK	780	
Qy	677	ELGMLKSIPLTLHSVSELVRLCRRSDVQBESEGSDDTDNKOSAAAFEDNEVQDEFLE--	734	
Db	781	ERQER-----EENER--RAKEVLQAE---NERKLEALEQENERRRLKETR	822	
Qy	735	-----KLETSEFFELTSEKQLILTALCHRLIMTYSVQDHMETTQQMSAELMWERLAV-L	788	
Db	823	EKENKKLREALELEKEK-----RLIEAF---ERAETERRLUKEDLEQEBEMRML	870	
Qy	789	KEENDKKRAEKOKRKEAKNGKENGKVENGLKTKDRKKR-IVKPEPQVDTAEADMISAVK	847	
Db	871	QBAKERLRHNRQHQENKQKH---EYSGEESDEKERDACAENEKTCETTYKE--AHGEQ	925	
Qy	848	SRRLLAIOAKKEREIQEREMKVKLERQAEERIRKHAAAF-----KAFQEGIAKAKLVMR	903	
Db	926	SSNESISDTLEENESIDNDVSNVKQKEEGTRQRESMSAETCPWKVFETKLKXDAQKEG	985	
Qy	904	RTPIGTDNRNNHRYWLFSDVEVPLFTEKGWVHDSIDYRNFHKKDHTVSGDEYDCPSRKA	963	
Db	986	TNEMDAD---TRLPERNEETP-----	1003	
Qy	964	NLGKNASMTQHGTATEVAVETTPKQGNLWFLCDSQK-----ELDELLNCLHPQIGRES	1019	
Db	1004	RLGEGNGCQNGQNGSEGEESTSVTENIIGKLI-----EOKSKNSTSKDASVLK-----RVS	1054	
Qy	1020	QLKERLEKRYQDIIHSHLARKPNLGLKSCDGNQLLNLFLRSDLIEVATRLQKGLGYVE	1079	

Db 1055 GLKTEVEERLEDDVVGVRDRQNPB---ESKSAPKTSYGF-RNHEYKFTHQOERGNITYEQ 1110
Qy 1080 ETSEPEARV-----ISLEKLUKDFGECVIALQASVIKFLQGFMAPKOKR-RKLQSEDS 1131
Db 1111 AGLNQDAKVERPLPSRVSVQREKE-----AERUKRERDUEMEOUL 1149
Qy 1132 AKTEEYDEKKMVEEAKVASALEKWKTAIREAQTSRMRHVLLGMLDACIKWMSAENARC 1191
Db 1150 RKVEEREREREREKDRMA-----PDQR---ALADARERLEKACAEARE 1190
Qy 1192 KVCPPKGGEDDKLLDCENKAFHLFCRLPALYVDPGEWCQACQAPATARRNSGRNYTE 1251
Db 1191 KSLP-----DKLSM-----EARLRAEAVERATSEARDRAA 1222
Qy 1252 ESASESDEDESEDEEBEEDYEVAGLRRLRPRTIRG-KHSVIPPAAAR----- 1304
Db 1223 EKAA-----FEARERNERSVDKQSOSGFFERMEISLDQFQNSVSGFASRVQDSHG 1277
Qy 1305 -SGRRPKPHSTRSQKAPPVDDAEDVDELVLQTKRSTRRSOSLELOKCEIILHKIVKYR 1363
Db 1278 TEGESQRTSLRLEHRTADRVIAET--LDTEVKRWSSGKEGNIRALLSTLQYLGP 1335
Qy 1364 FSW---PPREPVRDEADYDVIT---HPMDFQ-----TVQNKCS 1399
Db 1336 SGWQPLPLTEVITSAVKRAYRKATLCVHPDKLQORGANIHQYIC 1381

RESULT 35
D96796
probable heat shock protein, 53413-59028 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004
C:Accession: D96796
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Dewar, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: D96796
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-1871 <STO>
A:Cross-references: UNIPROT:Q9SRD5; GB:AE005173; NID:G6143906; PIDN:AAF04452.1; GSPDB:GN
C:Genetics:
A:Gene: F28016.15
A:Map position: 1

Query Match 3.3%; Score 265.5; DB 2; Length 1871;
Best Local Similarity 18.5%; Pred. No. 0.00028;
Matches 291; Conservative 220; Mismatches 581; Indels 483; Gaps 63;
Qy 33 RTREYEAERLEYSERIWTCK--STGSSQLTHKEAWEEOEVAELLKEEFPAYWEKL---- 87
Db 432 KDRQENDVGAQSEDISITKLQETGEQQFQGRKDKQENIKELREGOASAEAKNKND 491
Qy 88 VLEVHHNTASLEKLVDTAWLEIMTKYAVG-BECDFEVGKEMLVKVIKHPLEKVDDE 146
Db 492 ILKPVQKRGEGHKIKQT-FQETNKQPEGYNKIMETGK-----KINED 535
Qy 147 ATEKESD-----GACDSPS--SDKENSQ-----IAQDHQKKTWVKEDEGRRESINDRARR 196
Db 536 GTRKQVEMIRQOELDEPARSEKENSRRELVKSTNDEEKEKEIAGTE-RKEKESDR--- 591
Qy 197 SPRKPLTSUKGERKWAFFPHKYDVVKLQNECK-----IISNVPADSLI---RTERPPN 249
Db 592 -----PKILREQEVADEVAEDTKFSIYGEVKEEIEIAGKEFEKFGSD 633

Qy 250 KEIVRYF-----IRHNALRAAGTENAPWVEDELVKYSLPSKFSDFLLDPKYMTLNS 304
Db 634 DJIARIVRDTQELDSNAMK---QGE-----EKDMIQELVLEEK-----VCDGGKIATAVE 681
Qy 305 TIRKNTGSPDRKPSKSTNTSSLSPLNPKLWCHVHLKKSLSG-----SPL 351
Db 682 TTAENNKSRVQETBEQKDKEDT-----CGKHFOKLEIEGSDHGEVDEVEKGGK 732
Qy 352 KVKNSKNSKSPEHLEEM-----MKMSPNK-LHTNFI PKKGPAPAKP 394
Db 733 RTEAEKRIKDRAREABEIEKDLGVSGRYIKGTTIKELVENRGIYANEHEEKKODANRP 792
Qy 395 GRHSDK-----PLKAKGRSGILNGQS-----TGNSKS----- 423
Db 793 EKITGTIQELVLSNQLRQENVEDGDKTQELVEEKIKDCBEEEGSEESKIKTDDVVVKV 852
Qy 424 -----PKGGLKTPKTKMKQMTLMDMAKGTOK----- 449
Db 853 OGKIEELYKPKREHGTIKITELVEETGTGYEQEKEETAESIEACGSLRKVDGIEEHE 912
Qy 450 -----MTRAPRNSG-----GTPTSS 465
Db 913 LHEPKIHKERNRNVGAKFSGQEGKEEKIVESMTITENDNSIDVOETKKERGRLE 972
Qy 466 KPHKHLPPAALHLIAYYKENKDRKDRSALSVCVISTARLLSS----- 508
Db 973 SHDKYKIQELLMAGHNDKKEEQENNVTAEVELETERVSSKQVQEGKEDDNGSKFHE 1032
Qy 509 -EDARLPEELRSLVQRYELLEHKRWK-----SMSEORKEYL----- 547
Db 1033 PEERKSIEDTWTHEKREKRVLEEBETYPKDKHTGEDHNDHKEBEQENVIKAKAELNTE 1092
Qy 548 ---KKRRELKCK-----LKEKARERKEMLERLEKQKRYVEDQELT--GKNLPAPR 594
Db 1093 EDSFKVBEIEKQDHGELKRSVMVQAKROETEETKTRAMEKNETVERRKQTKDGS LGKUR 1152
Qy 595 LVDTPEGLPNTLFGDGVAMVVEPLSC-----YSGLLLPDAQYPITA-----VSLMEALSAD 644
Db 1153 EGEDFELGGHERRGEEDRIEELVEITEISDHKEKVKVKKDEDYILRSQDTGKVDGERERS 1212
Qy 645 KGGFLYLNRLVILLQTLQTLQDLQDEIADYGELGKMLSEIPLTLHSVSELVRLCLRRSD 704
Db 1213 K-----QRKHKSVDEIGDQEDADEAEEAAV-----VSRNENGSSRKVQ 1252
Qy 705 VQEESEGSDTDDNK--DSAAEDNEVDQEFLEKLTSEFFELTSEEKLIILALCHRLM 762
Db 1253 TIEESEKHKQNKIPETSNPEVNEDEERVVVEKETKEVAHVQE--LEGKTENCK---- 1306
Qy 763 TVSVQDHMETR-----QOMSAAE-LWKERL-----AVLKEENDKKRAEKKRKEAKN 809
Db 1307 ----DDGEGREERGKQGMATNMLRQRFKTKSDDGIVRKIQETKEEPEDEKKSQESS 1362
Qy 810 -----KENGKVENGLKGTDRK---KRIVKFPQVDTEADMISAVKSRLLA--IQAKK 858
Db 1363 HVVKLVADGSLRNLGELFSEKSTVSKMLKDESEKBEHKIRKPTERSNAPVIEKQG 1422
Qy 859 EREIOEREMVKLERQ-----AEEERIKHKAEEAKAFQGIKAK-- 899
Db 1423 NKNAAEBEQMDKIDRRGNKQNEIKGQEPYGVLRNGEHDKI TEYHREEKGTAEVNSTKIQ 1482
Qy 900 ----LVMRRTPIGTDNRNRYWLPFSEVPGLFIEBKGVWHDSDYRNFHCHDKDVTSGDE 954
Db 1483 QTKDELEKPRKPEISENHNHIEFMDSSQSODIEEG--SDQAE-KYAKONKIQEWMNDE 1539
Qy 955 D-----YCPRSKANLGNKSNMTOGTATEVAVETTPPKQGNLWF 996
Db 1540 DKKEBYHISERVNEMAKRIILQVESKANDGSSKKNTEQESTGLRGR----KREN--- 1592
Qy 997 LCDSOKELDELLNCLHPQIGRESOL---KERLEKYYQDIHSHILARKPNLGLKSCDGNQ 1053
Db 1593 ----HOELVELETSQKQKGVKDEYVVGAEIIEDEY-DSSRKIHHEERMSDKLEMHGEE 1647

C:Accession: A59252; B61231; G02055
 R:Phillips, C.L.; Yamakawa, K.; Adelstein, R.S.
 J. Muscle Res. Cell. Motil. 16, 379-389, 1995
 A:Title: Cloning of the cDNA encoding human nonmuscle myosin heavy chain-B and analysis
 A:Reference number: A59252; MUID:96025307; PMID:7499478
 A:Accession: A59252
 A:Status: not compared with conceptual translation
 A:Molecule type: mRNA
 A:Residues: 1-1976 <SIM>
 A:Cross-references: UNIPROT:P35580; GB:M69181; NID:G641957; PIDN:AAA91977.1; PID:G641958
 A:Experimental source: clone lib Lambda Zap II adult human T-cell library; cell line Jurkat
 A>Note: Between nucleotides 1942-1943 in mRNA encoding human brain MHC-B there is an alt
 R:Simons, M.; Wang, M.; McBride, O.W.; Kawamoto, S.; Yamakawa, K.; Gdula, D.; Adelstein, R.
 Circ. Res. 69, 530-539, 1991
 A:Title: Human nonmuscle myosin heavy chains are encoded by two genes located on different
 A:Reference number: A61231; MUID:91316803; PMID:1860190
 A:Accession: B61231
 A:Molecule type: mRNA
 A:Residues: 63-237, 'K', 239-664, 'L', 666-722 <S12>
 A:Cross-references: GB:M69181; NID:G641957
 R:Weit, L.
 submitted to the EMBL Data Library, August 1995
 A:Reference number: H00753
 A:Accession: G02055
 A:Cross-references: EMBL:U34304; NID:g1143217; PIDN:AAA84880.1; PID:g1143218
 A:Residues: 1-81 <WEI>
 C:Genetics:
 A:Gene: GDB:MYH10
 A:Cross-references: GDB:127350; GDB:G00-127-350; OMIM:160776
 A:Map position: 17p13-17p13
 C:Superfamily: myosin heavy chain; myosin motor domain homology
 C:Keywords: actin binding; ATP; coiled coil; hydrolase; methylated amino acid; nucleotide
 F:88-771/Domain: myosin motor domain homology <MMO>
 F:178-185/Region: nucleotide-binding motif A (p-loop)
 F:559-572/Region: actin binding #status predicted
 F:633-647/Region: actin binding #status predicted
 F:129/Modif site: actin binding #status predicted
 F:184/Binding site: ATP (lys) #status predicted
 F:701.711/Active site: Cys #status predicted

Query Match 3.3%; Score 263; DB 2; Length 1976;

Best Local Similarity 18.9%; Pred. No. 0.00039; Mismatches 566; Indels 610; Gaps 81;

Matches 335; Conservative 263; Mismatches 566; Indels 610; Gaps 81;

QY	15	PLPGEPPFTIHTQAFR----	TREYEALR-----	ERYSERIWTCKS	54
DB	316	PIPGQDKNFQETMEAHIMGFSHEILSMLKVYSSVLQFGNISFKERNTDQASMPEN	375		
QY	55	TGSSQLTH-----	KEAWBEO----	EVAELLKEFPANWYK	86
DB	376	TVAQKLCHLLGMNMFETRAILTPIKVGVDYVQKAQTEQADFAVEALAK----	ATYER	431	
QY	87	LVLEWVHNNTASLEK-----	LDVTAWLEIMTKYAVGECDEPVGCKRWLVKIVK	136	
DB	432	LFRLWLHINKALDRTRKQAGAFIGILDAGFEI-----	FELNSFEOL-----	474	
QY	137	IHPLEKVDDEATEKKSADGACDPSDKENSQTQADHQKKEVTVKDEGREGRESINDRARR	196		
DB	475	-----	CINYNEK-----	LQQLFNHTMFLBQEEVQREGI-----	504
QY	197	SPRKLPSTSLKGERKWAPKFEI	PHKYDVVKLQNEDKIISNVPADSLIRTRPNKKEIVRYF	256	
DB	505	-----	EW--NFIDFGLDLQ-----	PCIDILT--ERPANPPGVIAL	534
QY	257	IRHNALRACTGENAPWVE-----	DELVKVYSLPSKF-----	SDFLLDPVKYMT	300
DB	535	L-----	DEECWFPKATDKTFVEKLVQSGSHSKFKQPRQLKADFCIIHY-----	580	
QY	301	LNPSTKRKTGSPDRKPS-----	KSKTDNSSLSSPLN-----	PKLWCHVH-----	341
DB	581	-----	AGKVDYKADWLMKNMDPLNDNVATLLHQSSDRFAELWKNVDVRIVGLDQV	631	

QY	342	--LKSLSGSPKVKNSKNSKSPBEHLEMMKMS--	NKLHTNF----	HIPKGPAPKPKG	395
DB	632	TMGTMTAFGSAKYTKKGMFRFTVGVLKSLTKMLATLNTNPNFVRCIIPNH----	EKRAG	688	
QY	396	KHSDKPLKAGRSKGLNG-----	OKSTGNSKSPKGLTKPKTKMKQMTLLDMAGK	446	
DB	689	KLDPHLVLDQRCNGVLEGIRICROGFNRFVQEFRRYELITPNAIPK--	GFWDGKQA	746	
QY	447	TQKTRAPRNSGGTPTRTSPKPHKLPPAALHLIAYYKKNKOREDKRSALSVCISKAR--	L	505	
DB	747	CERMIRALELDPNLVRIQOSKIFFRAGVLAHL-----	EEERDLKITDIIIFQAVCRGC	800	
QY	506	LSSEDRARLPPELRSL--	VQKR--	YELLEHKRWASMSSEBQKRYLKKREELKCLKE	560
DB	801	LARKAFKKQOQLSALKVLQRNCAAYLKLHWQWVRVFTKVKPLQLQVQBEELQAKDEE	860		
QY	561	KAKERREKEMLE--	RLEKQKRYEDQELTGKLPAPFLVDTPPEGLPNTLFGDVAMVVEFLSC	619	
DB	861	LLKVKETQKVEGELEEMERKHQOLLEEKNI-----	891		
QY	620	YSGLLLPDAQVPITAVSLMEALSADKGGFLYLNRLVILLQTLLOLQDEIAEYGBELG	679		
DB	892	-----	LAEQLQA-----	ETELFAEAEMRRLA	914
QY	680	MKLSIPLTLHSVSELVRLCLRRSDVQSESGS----	DTDDNKDSAAFEDNEVQ--DE----	731	
DB	915	AKQOLEBILHDL-----	ESRVEEERNOILQNEKKKQMAHIQDLEBQLEDEEGA	965	
QY	732	-----	FLEKLETSEPPELTSEKQLQIL-----	TALCHRIIMTVSQVDMETROOM	776
DB	966	RQKLOLEKV--	TAAKIKKWEHEILLLEDQNSKFIKEKKLMEDRIAECSSQLAEEBEKAKN	1024	
QY	777	SALWKEKRLAVLKEENDKRAEKQKRRKEME--	AKNKENGKVENGLKTKRKRIVKPEQV	835	
DB	1025	LAKIRNKQVMISDLERLKKKEKTRQLEKAKRKLDCET-----	TDLQDQIAELQAI	1078	
QY	836	DTEADMISAVSKRLLAIO--	AKKEREIQ-----	EREMKVKLEQAE-----	877
DB	1079	DE-----	LKQLAKKEELQALARGDDETLHKNNALKVVRLEQAIQIAELQ	1124	
QY	878	ERIRKHAAAEKAFQEGIAKAKLVNMRTPIGTDRNHNRYLWFLSDVPGLFIKGGWVHDSI	937		
DB	1125	EDFESEKASRNKAEQO----	KDIL-----	SELEAKTE--	LEDTL
QY	938	DYRFNHCKDHTVSGDEDYCPRSK-----	ANLGN-----	ASMTQHTGATEV	981
DB	1160	D-----	TAAQOEL--	RTKREQVAELKKALBEETKNHEAQIQDMRQRHATALEE	1207
QY	982	AVETTP-----	KQGNLWFLCDSQKELDELNCLHPQGIRESOLKEREKRYQ	1030	
DB	1208	LSEQLQAKRFKANLEKNKQ-----	LETDNKLACVKKVLOQVKAESHEKRRKKLDAQVQ	1262	
QY	1031	DIHSIHLARPNLKGKCDGNOELLNFRSLIEVATRLQ-----	KGGLG	1076	
DB	1263	E-----	LHAKVSEGBD-----	LRVELAEKASKLQNELDNVSTLLBEAEKKGKIK	1305
QY	1077	YVEETSEFEARVISLEKLDKDFEGCVIALQASVKKFLOGFMAPKRRKRLQSEDSAKTEE	1136		
DB	1306	PAKDAASLESQLODQOEL-----	LOEETRQK-----	LNLSSRIQRLEBEKNSLQK	1351
QY	1137	VDEEKWVEAKVASALEKWKTAIRE--	AQTFSRMHVLLGLMDACIKWDMSEAKENACKVC	1194	
DB	1352	QEE-----	BEAR--	KXLEKQVIALQSQADTKKQVDDDLGTIE-----	SLEAKKLL
QY	1195	PKKGE-----	DDKILIDECNKAPHLFCLRPALYVEP--	DGEWQCPA-----	COP
DB	1399	KDAALSORLEEKALAYDKLEKTKR--	LOELDDLTVDLDHQRQVASNLEKKOKKFPQ	1455	
QY	1238	ATARNRSGRNYTESASEDSEDDSEDEEBE-----	EEEEEEEDYEVAGLRRLP--	1288	
DB	1456	LIAEKSISARYAE-----	RDRAEAAREKETKALSARALEBALEAKSEFEQRNQQLRAD	1512	

QY 1289 -RTIRGKSHVIPAARRGRRPGKPHSTRSOPKAPPVDDAEVDELVLQTKSSRRQSL 1347
Db 1513 MEDLMSSKDDV-----GKNVHELEKSKRAL-----EQVVEEMRTQLEELD----- 1553
QY 1348 ELQKCE-----EILHKIVKYRFSWPPREPVRDEAD-----YYDVITHPMDFOV 1393
Db 1554 ELQATEDAKLARLEVNQAMKAQFE---RDLQTRDEQNEEKRLILIKQVRELEAEDEDERK 1610
QY 1394 QNKCSCGYSRVOEFLTMKQVFTNAEVYNCRSHVLSM----- 1433
Db 1611 QRALAVASKKME---IDLKLEAQIEAANKARDEVIKQLKLAQAMQKDYQRELEBARAS 1667
QY 1434 -----VKTEQCLVLLHKLPHGVPRKRKPPDRLAEDGDESPAVGQ 1479
Db 1668 RDEIFQKSESEKKLLEAEILLOLEELASSRRARHAEQERDELADEITNS---ASGK 1724
QY 1480 SR--DEDR--SREAEIQEWLQDTSLSAKINSK 1509
Db 1725 SALLDEKRRLEARIAQLEEELEEEQSNWELLNDR 1758
RESULT 39
A40997
myosin heavy chain, striated adductor muscle - scallop (Aequipecten irradians)
N:Contains: myosin ATPase (EC 3.6.4.1)
C:Species: Aequipecten irradians
C>Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 09-Jul-2004
C:Accession: A40997; S13557
R: Nucleotide binding motif A (P-loop)
J. Biol. Chem. 266, 18469-18476, 1991
A:Title: Complete primary structure of a scallop striated muscle myosin heavy chain. Seq
A:Reference number: A40997; MUID:92011595; PMID:1917970
A:Molecule type: mRNA
A:Residues: 1-1938 <NYI>
A:Cross-references: UNIPROT:P24733; GB:X55714; NID:G5611; PIDN:CAA39247.1; PID:G5612
C:Superfamily: myosin heavy chain; myosin motor domain homology
C:Keywords: actin binding; ATP; coiled coil; hydrolyase; muscle contraction; nucleotide b
F:86-763/Domain: myosin motor domain homology <MMOT>
F:176-183/Region: nucleotide-binding motif A (P-loop)
F:547-586/Region: actin binding #status predicted
F:653-675/Region: actin binding #status predicted
F:836-1938/Domain: coiled coil #status predicted <COI>
F:836-1276/Region: S2
F:1277-1938/Region: light meromyosin
F:182/Binding site: ATP (Lys) #status predicted
F:693,703/Active site: Cys #status predicted
Query Match 3.3%; Score 262.5; DB 1; Length 1938;
Best Local Similarity 18.1%; Pred. No. 0.0004;
Matches 289; Conservative 240; Mismatches 561; Indels 511; Gaps 67;
QY 90 EMVHNTASLE---KLVDTALEIMTKYAVGECDFEVQKEKMLKVIK- IHPLEKV 143
Db 486 QFNHMFLEQBEYKKEGIAW-----EFIDF--GMQLQMCIDILIEKPMGLSIL 533
QY 144 DEEAETKSGDAGDSPSKENSQIAQHQKETTIVVKEDEGRRESINDRARRSPKPLPT 203
Db 534 EEECMFPKAD-----DKSFQKLYQNHMK-----NRMFTKPK-PT 569
QY 204 SLKGRKRWAPKPLPHKYVDVQLQNEDKIISNV-ADSLIRTEPPNKEIVRYPIRN 260
Db 570 RPNQ-----PAHFELHHY-----AGNVPYSITGWLEKNKDPINENVV----- 607
QY 261 ALRAGTGENAPVVEDELVKYSLPSKFSDFLLDPKYMTLNPSTKNTGSPDRKPSKK 320
Db 608 ALLGASKE--PLVA--ELFKAPEEPA-----GGGKKGKGS 639
QY 321 -----SKTDNSLSPLNPKLWCHVHLKLSGSPKVKNSKNSPBEHLEENMMQMS 374
Db 640 SAFQTISAVHRSNLKMLKMLVSTPHFVRCIIPNELKQPLVDA-----ELVHLQIQ 692
QY 375 PNKLHTNFHIPKGGPPAK---RFGKHSKPLKAKGRSGKILNGKQSGTGNKSPKGLKTP 431

Db 693 CNGVLEGIIRICRGPFPSRLIYSEFKQYSILAPNAIPQGFVDG-----KTV 738
QY 432 KTKMKQMTLLDMAK---GTQKMTAPRNSGGTPT--TSSKPHKLPAAALHLIAY--KEN 485
Db 739 SEKILAGLOVDPAYRLGTTTKVFFKAGVGLONLEEMDERLSKIISHFOAHIRGYLTRKAY 798
QY 486 KOREDKRSALSCVISKTARLLSSSEDRARLPEELRSVLQKRYELLEHKKRWASSEEQRKE 545
Db 799 KKLQDQKIGLS-VIQRNIR-----KWLVLNRWQMKLYSKVKPPL 837
QY 546 YLKKKKEELKKLEKAKERREKEMLERLEKQKRYEDQBLTGKVLPAFLVDTPEGLPNT 605
Db 838 STARQEEEMKEQLQMDKMKEDLAKTERIKKEEQNVTLLEQKNDLFLQLOTL-----D 893
QY 606 LFGDVAMVVEFLSCYSGLLLPDAQYPTITAVSLMEAL-----SADKGGFLYLNRLVILL 659
Db 894 SMGDOEERV-----KLIMQKADPESQIKELERLLDEEDAAADLEG----- 935
QY 660 QTLLOTLQDETAEDYGEGLMKLSEIPLTLHVSSELVRLCLRRSDYQEESESGSDTDNDK 719
Db 936 -----IKKMEADNANLKKDIDLENTL-----QKAEQDKAHKDNQI 972
QY 720 SAAFEDNEVQDELEKL-----ETSEFFELTSEBKLIULTALCHRLMTY--SV 766
Db 973 STLQGEISQDDEHIGKLNKEKKALEBANKTSDSLQ-AEEDKCNHLNKLKAKLEQALDEL 1031
QY 767 QDHMETROQMSAELWKERLAV--LK--EEN--DKKRAEKQKKEAKNKE-----NGKV 815
Db 1032 ENLREKRVGDEKAKKVKQDLKSTOENVEDLVRKRELEENVRRAEISSLSK 1091
QY 816 E---NGLKGTDRK-----KRVKFPQVDTEADMI SAVKSRRLAIQAKGREIOE--- 864
Db 1092 EDEQNLVSQKIKELQALQARIEELEEELEARNARAKVEKQ-----AELNREBELGE 1145
QY 865 -----REMKVLEQRAEERIKHKAASAKAFQEGIAKLVVRRPTIGTDRNHR 916
Db 1146 RLDEAGGATSAQIELNKKREAEELLKIRDLLEASLQHEAQISALRKHKQDAANE----- 1199
QY 917 WLFSDEVPLGFTKEGVH-----DSIDYRNHHCKDHTVSGDEYDPRSKKANLG 966
Db 1200 ---MAQVDOLQKVKLEKDKDKDLKREMDLDSQTHNMKNKGS-----EKVMQKQFS 1251
QY 967 KNASMTQHTGTA TEVAVETTPKQGNLWFLQDSQKELDELNCLHPOGIRSQLERLE 1026
Db 1252 QMSDLNAR-----LEDSQRSINEL-----QSQ-KSRQ 1278
QY 1027 KRYQDIHSHILARKPNLGLKSCDGNQELLNFLRSLDIEVATRLQKGLGYVEET---SE 1083
Db 1279 AENSDLTROLEDA-----EHRVSVLSKEKSQLSQLEDARRSLEETTRASK 1325
QY 1084 FEARVISL-----EKLKDFGECVIALQASVIKKFLOGFMAPKQKREKLOSEDSAKTE 1135
Db 1326 LQNEVRNHHADMTAREQLEEEQES-----KSDVQRLSKANNEIQOWRSKFSEGANRTE 1381
QY 1136 EYDEEKKVMVEAKVASALEKWKMTAIREAQTFSRMHVLLGMLDACIKWDMASENARCKVCP 1195
Db 1382 ELEDQKRC-----LIGKLESEA-EQTTEANAKCSALE 1412
QY 1196 KKGCE-----DDKLILDCENKAFHLFCRLPALYPVDPGEWQCPA-----CQPATARNR 1245
Db 1413 KAKSRLOQLEDMSEIVDRANASVQMEKKQAFDKTTAEMQAKVNSLQSELENSQKESR 1472
QY 1246 GRN---YTEESASEDEDD-----EVDLVLQTKSSRRQSLQLEQKCEBILHKV 1362
Db 1473 GYSALYRIKASIEEYQDSIGALRENNKNLADEIHLDTQLDQSEGRSTHELDKARRRLEM 1532
QY 1264 DEEE-----EEEEEEEDYEAGLRRLPRKTIRGKSHVTPPAARSGRRPK 1311
Db 1533 EKEELOAALEAEAGLEBEAKVMPRAQLEIATVRNEIDKRIQEKKEEFNDTRNRHQRAL 1592
QY 1312 KPHSTRSQPKAPPVDDA-----EVDLVLQTKSSRRQSLQLEQKCEBILHKV 1362

Db 1593 SMOASLEAEAKGKA--DAMRIKKKLEQDINELEVALDASNR-----GKAE--MEKTVK- 1641

Qy 1363 RSNWPREPVT-----RDAEDYYVITHPMDPQTQVQKSCGSRVSVOEFLTDMKQ 1414

Db 1642 RYQOQIREMOTSIBEQORDARESYNN-----AERCTILMS-GEVEELRAALEQ 1691

Qy 1415 VFTNAEVNCRGSHVLSVMVKTEQCLVLLHKHLPQHFPVVRKRKKFPDRLAEDGSDSEP 1474

Db 1692 -----AERKASONELEA-DANDRVN 1711

Qy 1475 EAVGO--SRDEDRRGREABIEQWLQDTSLYSAKINSKDHNC 1513

Db 1712 ELTSQVSSVQGGKRLKLEGDINAWQTDLDHMGELGADERC 1752

RESULT 40

A:7297

C:Species: Xenopus laevis (African clawed frog)

C>Date: 22-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 09-Jul-2004

C:Accession: A47297; A55441

R:Bhatia-Dey, N.; Adelstein, R.S.; Dawid, I.B.

P:Proc. Natl. Acad. Sci. U.S.A. 90, 2856-2859, 1993

A:Title: Cloning of the cDNA encoding a myosin heavy chain B isoform of Xenopus nonmusculi

A:Reference number: A47297; MUID:93219383; PMID:8464900

A:Accession: A47297

A>Status: preliminary; not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 1-1992 <BHA>

A:Cross-references: UNIPROT:004834; GB:I09740; NID:G214623; PIDN:AAA49915.1; PID:G214624

A:Experimental source: XTC cells

A>Note: sequence extracted from NCBI backbone (NCBIP:128722)

R:Kelley, C.A.; Oberman, F.; Yisraeli, J.K.; Adelstein, R.S.

J. Biol. Chem. 270, 1395-1401, 1995

A:Title: A Xenopus nonmuscle myosin heavy chain isoform is phosphorylated by cyclin-p34

A:Reference number: A55441; MUID:95138137; PMID:7836406

A:Accession: A55441

A>Status: preliminary

A:Molecule type: protein

A:Residues: 198-232 <REL>

C:Superfamily: myosin heavy chain; myosin motor domain homology

C:Keywords: nucleotide binding; P-loop

F:88-787/Domain: myosin motor domain homology <MMOT>

F:178-185/Region: nucleotide-binding motif A (P-loop)

Query Match 3.3%; Score 262.5; DB 2; Length 1992;

Best Local Similarity 19.1%; Pred. No. 0.00041;

Matches 220; Conservative 207; Mismatches 453; Indels 269; Gaps 44;

Qy 39 EARLERYSERIWTCKSTGSSQLTHKEAWEE--QVEAELLKEFPANWEKLVLEVMHNT 96

Db 992 EAKIKKMEEDILVLEDQNSKFLKEKKLLEERIAESTSQAEEKA---KNLAKLNKQE 1048

Qy 97 ASLEKLVDTAWEIMTKAV-----GEEDFVVG-KEKMLKVIVKIHPLEKVBEEA 147

Db 1049 WMISLEERLKEEKTRQELEKAKKLDGETTFDQDAELQAQIEELKQ-LAKEBEL 1107

Qy 148 TEKSDGACDSPSSKENSQIAQDHQKKEITYVVEDEGRESINDRARRSPKLTSLKK 207

Db 1108 QAALARG--DEEVLOKQNTLKLVLRELQAQIAELQEDLESEKASRNKAQEKQKDLSELEA 1165

Qy 208 GERKWAPEKFLPHKVDVLQNEDEKIIISNVPAQSLRTERPPNKEIVRYPIRINAIKAGTG 267

Db 1166 -----LKTELED-TLDTTAAQQLRTKREQVABLRSKISEET----- 1202

Qy 268 ENAPWVVEDELVKYSLSPKSFDFLLDPKYVMTLPNTPSKTKNTGSPDRKPSKSKTDNS 327

Db 1203 RNHEAQIQEMRQRTALEESEL-----EQAKRFVNLEKKNQKQSL--DNKE 1250

Qy 328 LSSPLNPKLWCHVHLKATSSGPLKVNKSNKSPPEELHEM-MKQMSPNKLTFTNPHIPK 386

Db 1251 LATEV-----KSL--QOMKAESYKRRKKLEGGVQELHAKVLEGRADW---- 1293

Qy 387 KGPPAKKPGKSHDKPLKAKGRSKGILNGKQKSTGNSKSPKKGKLTPTK-TMKQMOTLDMAK 445

Db 1294 -----VEKSKLQNELEV-----SSLLEAEKKGIGLAKOVASHESQLQDTQE 1337

Qy 446 GTQKMTAPRNSGGTPTTSKPHKLPAPALHLIAYYYKXKDKREDKRKLSALSCVISKARL 505

Db 1338 LLQEBTRQKLAQSSRIQLEBKNNLQE-----QOEEREEARKSLEKQILSLQSQL 1388

Qy 506 LSSDRARLPEB---LRSLVQKRYELLEHKRWASMSSEQRKYLLKKREELKKKLKEKA 562

Db 1389 I--EAKKVVDDVGTGIELEEVKKLLKDTGLGQRLBEKIIAY--EKLEKTKNRLQOEL 1444

Qy 563 KE-----RREKEMLERLEKQKRYEDOELT-GKNLPAPFLV--DTPEGLPNTLFGDVAMV 614

Db 1445 DDLVMDLHQIQTIVSNLEKKOKKFDLLAEKKNISARHAEERDRAE----- 1491

Qy 615 EFLSCYSGLLLPDAQYPT-----AVSLMEALSADKGGFLYLNRLVLILLQTLQTLQD 669

Db 1492 -----DAREKETKALSARALDEALEA-QDEFERLNKQLRAEMEDLMSS--KD 1536

Qy 670 EIAEDYGBELGKMLSEIPLTLHSVSELVRLRRSDVQSESEGSDDTKNKSAAFPEDNEVQ 729

Db 1537 DVGKNVHEL-----EKSRRALDQOVVEEMRTQLEELE 1567

Qy 730 DEFLEKLETSFFELTSEKLIQILTALCHRIILMTYSVQDHMETR-----QQMSAELMKE 783

Db 1568 DE-----LQGTEDAKLRLVNMQAMKAQPERDLQTRDEQNEEKRALVQKVRLEAELEDE 1623

Qy 784 R-----LAVLKE-ENDKKRAEKQKRKEMEAKNKGKGVNGLGKTDKTKRIRVKEPQVD 836

Db 1624 RKQRAWAVAIKKLEMDMKDFESQ-----IEAANK-----GREDAIKQLRKLQAQTK 1670

Qy 837 TEAEDMISAVKSRLLAIQAK-EKEIO--BREMVKVLEROQAEERIRKHK----- 884

Db 1671 DYQRELEEARASRDDIFAQSKENKELKGLAEAILQOEELAAASRSRSHABQERDELAD 1730

Qy 885 -----AAAEKAFQEGIAKALV--MRTPITGTD----- 910

Db 1731 EISNSTSGKSALLDEKRLREARIAHLEEBEESQNMELNDRFKTILQVDTLNSELAA 1790

Qy 911 -----RNHRYWLFSDVPGFLIEKGWVHDSIDYRFNHCKDHTVSGDEDYCPRSKXAN 964

Db 1791 ERSQSKSENARQQLERQNKELKAKLQELGSKVSKPK-----ATIAATLE-----SKIAQ 1840

Qy 965 LKQNASMTQGTATEVAVETTPKQGNLWFLCDSQKELDELLNCLHPQIGRESQLKER 1024

Db 1841 LEEQLEQEAKEVASNKLVRTEKKLKEVFMQVEDERHADQYKEQMEKANTRMKQLRKQ 1900

Qy 1025 LEKRYQDIIHSTHLARKPNLGLKSCDGNQELLNFLRSLIEVATRLQKGG-LGYVEETSE 1083

Db 1901 LEEAEEETRANASARKLO---RELDDATEANEVLRSRVSTLKNRLRGGPVSPFSSSSSR 1957

Qy 1084 FEARVISLE 1092

Db 1958 SGRRQLQIE 1966

RESULT 41

A54514

Glutamic acid-rich protein precursor - malaria parasite (Plasmodium falciparum)

N:Alternate names: GARP

C:Species: Plasmodium falciparum

C>Date: 15-Oct-1994 #sequence_revision 15-Oct-1994 #text_change 09-Jul-2004

C:Accession: A54514

R:Triglia, T.; Stahl, H.D.; Crewther, P.E.; Silva, A.; Anders, R.F.; Kemp, D.J.

Mol. Biochem. Parasitol. 31, 199-202, 1988

A:Title: Structure of a Plasmodium falciparum gene that encodes a glutamic acid-rich pro

A:Reference number: A54514; MUID:89040048; PMID:2903445

A:Accession: A54514

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-678 <TRI>

A:Cross-references: UNIPROT:P13816; GB:J03998; NID:G160298; PIDN:AAA29605.1; PID:G160299

C:Genetics:

A:Introns: 25/3

C:Superfamily: histone H1

C:Keywords: tandem repeat

Query Match 3.3%; Score 261.5; DB 2; Length 678;
Best Local Similarity 19.2%; Pred. No. 0.00013;
Matches 159; Conservative 126; Mismatches 294; Indels 247; Gaps 24;

QY 483 KENKREDKRSALSCVISKTA-----RLLSSEDRARPEELSLVQRYELLEHKKRWASW 538
DB 70 KEEKDEKDDVPTISNDNLKNAHNNFISSTD-----PTNIIIVNDKDNENSVDKKK--DK 123

QY 539 SEQRKEYLKKREELKKLKEKAKERREKEMLERLEKQRYEDQEL-----TGNLPAP 593
DB 124 KKK 183

QY 594 RLVDTPGELPNTLFGDVAMVVEFLSCYGLLLPDAQYPTAVSLMBALSGGFLYLR 653
DB 184 ATEHGEENL-----DEEMVSEINNAQGLLLSSPYQY-----REQGG----- 221

QY 654 VLVIILLQTLQLLODEIAEDYGLGMLKSEIPLTLHSVSELVRLCLRSDDQESGSD 713
DB 222 -----CGIISVHETS--ND 234

QY 714 TDNDKDSAAFEDNEVQDFLEKLETSFPFELTSEELQIILTALCHRLMTYSVQDHMETR 773
DB 235 TXDNCKENISDEKEDHQOEMLKTLDDKKERQKEK-----EMKEQEKIEKK 281

QY 774 QMSAELWELRALVKEENDKGAERQKQKEMAKKNGKGVNGLGKTDKRIIVKFPF 833
DB 282 KKKQEEKKK-----KQEKERKKQEKERQKQKEMKQKQKIEKKERKKKEKKKKKKHD 336

QY 834 QVDTEA--EDMISAVKSRRLAIQAKEREIOEREMVKLRQAEERIRKHAAEKA 890
DB 337 KENEETMQPDQTSSETNNEIWPPLSPDVTPTPEHKEGEHEKEHEGEHEKEHEKE 396

QY 891 FQGIKAKLVMRRTPIGDRNHNRYWLFSDVPGLFIEKGWVHDSIDYRPNHCKDHTV 950
DB 397 EEH-----KEEHHKE-----EHSKEHKS 416

QY 951 SGDEDYCPRSKKNALGNASMTQHTATEVAVETTPPKQGNLWFLCDSQKELDELLNC 1010
DB 417 KGKCD-----KGKDKGK-----HKKAKKEKVKHVVKN-----VIEDEDQGVIEIIL 460

QY 1011 LHPQGIRESQLERLEKRYQDIHSHLARKNLGLKSC---DGNQELNLFRLSDILEVA 1067
DB 461 -----EDKEACEQH-----ITVESRP--LSQPOCKLIDPEPQLTMDKSKVEKN 504

QY 1068 TRQKGLGYVBETSFEARVISLEKLDGFCVIALQASVIKKFLQGFMAPKQRRKIQ 1127
DB 505 LSIQEQOLICTIGRVNVVPRDNHKKQWAKIEAEALQKQKVDKE-----EDKESSEVQ 559

QY 1128 -----SEDSAKTEEVDEBKQWVEAKVASLEKWTATREAOFTSRMHVLLGMDACIK 1181
DB 560 ERSKEVQDEEEVEDE 602

QY 1182 WMSAENARKCVPKKGEDDKLLILDCENKAFILFCLRPALVEVPDGEWQCACQATAR 1241
DB 603 -BEDEDEDEDDAEDEDD-----AE 623

QY 1242 RNSGRNRYTESASESDDE 1287
DB 624 EDEDDAEEDDDDEDDDE 669

RESULT 42

T51505

hypothetical protein F5E19.70 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 18-Aug-2000 #sequence_revision 18-Aug-2000 #text_change 09-Jul-2004

C:Accession: T51505

R:Sato, S.; Nakamura, Y.; Kaneko, T.; Kato, T.; Asamizu, E.; Kotani, H.; Tabata, S.; Mew

submitted to the Protein Sequence Database, August 2000

A:Reference number: Z25394

A:Accession: T51505

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-853 <SAT>

A:Cross-references: UNIPROT:Q9LFE4; EMBL:AL391147

A:Experimental source: cultivar Columbia; BAC clone F5E19

C:Genetics:

A:Map position: 5

A:Introns: 6/2; 79/3

A:Note: F5E19_70

Query Match 3.3%; Score 261; DB 2; Length 853;
Best Local Similarity 19.7%; Pred. No. 0.00018;
Matches 195; Conservative 146; Mismatches 368; Indels 282; Gaps 38;

QY 435 MKQMTLLMAKCTQKMTAPRNSGGTTP-----TSKPKHHLPAPALHIIAYY 482
DB 1 MAAKTKTSLSETTTTTPTGKSSPATPRIAKRTVANKSETSNNSPSTTTTPHSRLSLDRSS 60

QY 483 KENKREDKRS-ALSCVISKTAARLLSSSEDRARPEELSLVQRYELLEHKKRWASWSEE 541
DB 61 PNSKSSVERRSPLPTPPPKSQARVAAVKGTSPOTTTLSQIKEDLKKANERISLEKD 120

QY 542 QRK-----EYLKKREELKKLKEKAKERREKEMLERLEKQRYE-----DQELTGK 588
DB 121 KAKALDELQAKKEAEQVTLKDDALKQAKHVEENSEIEKFOAVEAGIEAVQNNEBELKK 180

QY 589 NLPAPFLVDTPGELPNTLFGDVAMVVEFLSCYGLLLPDAQYPTAVSLMBALSGGFG 648
DB 181 EL-----ETVKNQHASDSA-----ALVAVRQEL 203

QY 649 LVLRNVLLVILLQTLQLLODEIAEDYGLGMLKSEIPLTLHSVSELVRLCLRRSDVQEE 708
DB 204 EKINELAAAFDAKSKALSQAEDASKTAIEHAKVDI-----LSSELTRL----- 248

QY 709 SEGSDTDNDKDSAAAFEDNE---VQDEFL---EKLETSEFFELTSEELQIILTALCHRL 761
DB 249 ---KALLDSTREKTAISDNEMWAKLEDEIVLKRDLJESARGFEAEVKEKEMIV---EKLN 302

QY 762 MTYSVQDHMETRQMSAELWELRALVKEENDKGAERQKQ-----KEMEANKE 811
DB 303 VDLEAAKMAESNAHSLSNWQSKAKELBEQLEE--ANKLERSASVLSVSMVKLEGSNDK 360

QY 812 NGKVENGLKTDKRIIVKFPQVDTEADMLSAVKSRLLAIQ---AKKREI----- 862
DB 361 LHDTEI--TDLKERIVTLETTVAKQKEDL--EVSQRLGSEVEEVSQNEKEVEKLKSE 416

QY 863 -----QEREMVKLRQEA-----EEERIRK-----HKA 885
DB 417 LETVKEKNRALKKEQDATSRVORLSEKSKLLSDLESSEKESKKAESLASALHEV 476

QY 886 AA-----EKAPQEG-----IAKALVMRRTPIGDRNHNRYWLFSDVPGLFIEKGW 932
DB 477 SSEGRELKEKLSQGDHEVETQIDDLKLVIKAT-----NEKYENMLDE----- 519

QY 933 VHDSDYDFN--HHCKDHTVSGDEDIYCPRSKANL-----GKNASMTQHG 976
DB 520 ARHEIDVLVSAVEQTKGHFESSKDW--EMKEANLVNYVKWMEEDVASMKGEMRDLNL 577

QY 977 TATEVAVETTTTPKQGNLWFLCDSQKELDE-----LLNCLHPQGIRESQKLERL---EKRY 1029
DB 578 KTEEEADAANKKEAQTCK---DSLKEVEEIEVYQETLGEAKASMKLKENLLDKETEF 633

QY 1030 QDIH-SIHLARKPNLGLKSCDGNQELNLFRLSDILEVATRLQKGLGYVEETSFEARV 1088
DB 634 QNVIHENEDLAKKEDVSLKKI---EELSKLLEEAIALAKKQPEENG-----ELSESEKY 685

QY 1089 ISLEKLDKFGECVIALQASVIKKFLQGFMAPKQKRLQSEDSAKTEEVDEBKQWVEAK 1148
DB 686 DLLPKVVEFS-----SENGHRSVEEKSAAKVTLDHPEF---POQO 721

A;Status: preliminary; translated from GB/EMBL/DBSJ
A;Molecule type: mRNA
A;Residues: 1-2442 <MAC>
A;Cross-references: EMBL:AF022655; NID:g2832236; PIDN:AAC06349.1; PID:g2832237
A;Experimental source: cell line HeLa

Query Match 3.3%; Score 260; DB 2; Length 2442;
Best Local Similarity 18.6%; Pred. No. 0.00068;
Matches 292; Conservative 259; Mismatches 608; Indels 412; Gaps 67;

QY 27 HTCEARTREVEEARLERSYRTWCKSTGSS-----OLTHKEAWEEBQVAEL 75
DB 851 HEKEVQLRKWE-----KERSWHQOELAKALESLEKREKMELEMLKQO-QTEMEAIQA 903
QY 76 LKEEFPWVE-----KLVLVHVHTASLEKLVDT-----AWLEIMTKYAVGECDFE 123
DB 904 QREERTQAESALCQWLETEKERVSLLETLLQTKELADASQQLERLQ-----DMK 956
QY 124 VGEKMLKVKIVKIHLEKVEDEATEKKSOGACDSPSSDKENSS--QIAQDHQK-----K 176
DB 957 VQKLKEQETTGILQTOLEAQRELKEAARQHRDDLAALQEBSSLLQDKMDLQKQVEDLK 1016
QY 177 ETVVKEDEGR-----ESINDPARSPKLTSLKKGKRWAPKFLPHKYDVKLQNEDKII 233
DB 1017 SOLVAQDDSORLVEQVQEKLRQET-QEYNRIKELERERKASLT-----SLMEKEQREL 1069
QY 234 SNVPADSLIRTPPPNKEIVRYFIRHNALRAGGENAPVVEDELVKYSLPSKFSDFLL 293
DB 1070 VLOEADSIROELSAIRQ-----DMQEAQGEQKELSAQVELLQO-EVEKEADFLA 1119
QY 294 DPYK-----YMTLPNTPTRKNTGSPDRPKSPKSKTD 324
DB 1120 QEAQLELELEASHITEQOLRASLWAQEAQAAQLHLRLSTESQLEALAAEQPGNQAQAQ 1179
QY 325 N--SSLSPPNPKLVCHVHLKXLSG-----SPUKVXNKNKSKSPHLEEMKQWSPNKL 378
DB 1180 AQLASLYSALQALGVCESRPELGGDSAP-----SVWGLEPDQ-----NGA 1223
QY 379 HTWFHPKGPAPKPKGSHDKPLKAGKSGILNGOKSTGNSKPKGKTKPTKWKQM 438
DB 1224 RSLF-----KRGPLL-----TALSAAVAVASALLKHLQDLWKTQOTRDVLRDOVQKLE-- 1270
QY 439 TLLDMAKGTQKMTAPRNSGTPRTSSKPHKHLPPAALHLIAYYKENKQREDKRSALSCV 498
DB 1271 -----ERLT-----DTEAKSQVHTELQDLQRLSQNSQEKSKWEGQNSLESE 1314
QY 499 ISKTARLLSS-EDRALPELRLSVQRYELLEHHKGRWASMEEQ-----542
DB 1315 LMELHETMASLQSLRRAELQRMQAQGERELLQAAKENLTAQVEHLQAAVVEARAQASAA 1374
QY 543 --RKEYLKKREELKKLEKAKEREKEMLERLEKQRYEDQELTGKILPAF--FLVDT 598
DB 1375 GILEEDLRTARSALKUNBEVERBRAQALQOGELKVAQGKALO-ENLALTQTLAER 1433
QY 599 PEGLPNTLFGDVAWVVEFLSCYGLLPDAQYPIITAVSLMEALSADKGGFLYLNRLVLIL 658
DB 1434 EEEV-ETLRQIQOEL-----EKQREWQKAAI-ELLSLD-----L 1465
QY 659 LQTLQTLQDEIAEYDGLGMLKLSRPLTLHVSSELVRLCLRRSDVQSESGSDTDDNK 718
DB 1466 KKNQEVLDQEQIQELEKCRSVLEHLPMAVQEREQ--KLTVQREQIIREPEKORETORNV 1523
QY 719 DSAAFEDNEVDQDFLE-----KLSTSFELTSEE---KLQILTAL-----756
DB 1524 LEHQLLELEKQDMIESQRQGVODLKQVLTKLALALEENHNHKECOQKILKELEGOR 1583
QY 757 -CHRIIMTVSDVDMHTROQMSA-----ELWKERLAVLKEENDKKRAEKQKKEKMEAXN 809
DB 1584 ETQRVALTHLTLDEERSQELQASSQIHDLSEHSTVLARELQEROEVKVSQREQIEELQ 1643
QY 810 KENGKRVENGKGTDRKKRIVKFPQVDTEADDMISAVKGRRLAIAQAKKER-----ETQERE 866
DB 1644 RQKEHLTQDLERRDQELMLQKRIQV---LEDQ---RTRQTKILEEDLEQIKLSUREG 1696

QY 867 MKVKLERQAEERIRKHKAAAEKAFQEGIAKAKLVMRTPITGDRNHNHYWLFSDVPG 926
DB 1697 RELTTQRLQMQRAEBEGKPS-KAQRGSLHEHMKLILR-----DKEKEV 1738
QY 927 FIEKGWHDSDIYRNHHCKD-----HTVSGDEYCPRSKKAAN---LGNASMTQ 974
DB 1739 ECQEQHIHEL-----QELKDQLEQOGLHRKVGETSLLLSQREOEIVLQOQAEARE 1792
QY 975 HGTAETAVAVETTPKQGNLWFLCDSQKEL---DELLNCL-----HPQGIRESQKLERLE 1026
DB 1793 QELKEQSLSQO-----LDEAQRALARDQDELEALQOEOQQAQO-QSERVKEKAD 1841
QY 1027 KRYQDIHSHIARKPNLGLKSCDGNQELNLFRLSDLEIETVATRLQ--KGLGYVE-ETSE 1083
DB 1842 -ALQGALEQAHTLKERHG--ELQDHKEQARRLEEBELAVEGRRVQALEEVLGDLRAESRE 1898
QY 1084 FEARVISLEKLDKFGECVIA-----LOASVIKK-----FLOGFMAPKQKRR 1124
DB 1899 QEKALLALQO-----QCAEQAEHEVETRALQDSWLQAQVLKQERDOELEALRAESQSSR 1953
QY 1125 KLOSEDSAKTEVDE-----EKKMVEEAKVASALEKWKTAIREAQTFPSRMHVL 1172
DB 1954 HQEEAARARAEALQALGAHAAALQKGOHLLLEQALSRSLEA-STATIQAS-----2004
QY 1173 LGMLDACIKWDMASENARCKVCPKGGEDDKLILCDECNKAFHLFCILRPALYEVDPGEWQ- 1231
DB 2005 ---LDACQAHSQLEEA-----LRIQEGEIQDQDLAY 2033
QY 1232 ---CPACQAPATARRNSRGNYTE-ESASEDSDESDSEEBEEEE-----EEBEEYEVAG 1283
DB 2034 QEDVQOQOALQARDEELRHQERELQLEKSLAQRVQENMIQEKQNLGLERBEE--EIRG 2091
QY 1284 LRLRPK---TTRGKHSVITPPAARSRRPGKK--PHSTERSQPKAPPVDDAEVDELVLQ 1338
DB 2092 LHQSRELOLTLAQEQEILELRETQQRNNLEALPHSHKTS-----PMEOSLKL 2141
QY 1339 KRSSRRQSLQLEKCEILHIVKIRFSPFPVTRDEAEYDYVITHPMDFTQVQNKCS 1398
DB 2142 DSLEPRLQRELERLQALRQTEAREIEW-----REKAQD-----LALSQAQTKAS 2186
QY 1399 CGSYRSVQBEFL 1409
DB 2187 VSSLQEVAMFL 2197

RESULT 47
T18372
repeat organellar protein - Plasmidium chabaudi
C:Species: Plasmidium chabaudi
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T18372
R:Werner, E.B.; Taylor, W.R.; Holder, A.A.
Mol. Biochem. Parasitol. 94, 185-196, 1998
A:Title: A Plasmidium chabaudi protein contains a repetitive region with a predicted sp
A:Reference number: Z18922; MUID:98418765; PMID:9747969
A:Accession: T18372
A:Status: preliminary; translated from GB/EMBL/DBSJ
A:Molecule type: DNA
A:Residues: 1-1939 <WER>
A:Cross-references: UNIPROT:Q25662; EMBL:U43145; NID:gl1151157; PID:gl1151158; PIDN:AAC63

Query Match 3.2%; Score 259.5; DB 2; Length 1939;
Best Local Similarity 18.1%; Pred. No. 0.00055;
Matches 264; Conservative 250; Mismatches 503; Indels 439; Gaps 61;

QY 36 EBYEARLERSYRITCKSTGSSQLTHKEAWEEBQVAELKEEFPWYKVLVEMVHN 95
DB 96 KEYEVKIELEKEL---KLEKEKQI-NKEYEKELNEKSEFIKQWELLKEKELINLKEN 151
QY 96 -----TASLEKLVDTAWLEIMTKYAVGECDFEVEGKEKMLKVKIV-----KIHPLEK 142
DB 152 KINNKEIITLKREKINDIESEYIEKNKEKELNVEVTNIKMSLDKLTCEVQEKKNLEK 211


```
QY 143 VDEATEKSGDCAKSPSSDKENSQIAQDHOKKTVVKKEDGRRESINDR----- 193
Db 212 INKKVIEKEN-----NLRELKEFMKEKNEIIESLDG-----TINDKKNAYEKUEI 256
QY 194 ----ARRSPKRLPTSLKGERKWAPPKFLPHKYDYVKLQWEDKIIISNVPADSLIRTERPPN 249
Db 257 SPEEKRMTEMLDLSKLIKKEENFA-----NKQAKLEKENIIEIKLKDIESREKDPKS 309
QY 250 KIVIRYFIRH--NALRAGTGENAPWV-----VEDELVKK-----YSLPSKFSDFLLD 294
Db 310 KEEKFASMENELNTLKSLSKNACQMEVYKLEIKDLSQSLVEKEREIFEIKNEYDKINN 369
QY 295 PYKYMTLPSTYRKWT--GSPDRKSKSKTQNSLSPLNFKLWCHVHLKXSLSGSLK 352
Db 370 MKEKUSSINDGIDNTVLHSEBEKINKLKEKETELN-----ETHKKNYLSIE--T 418
QY 353 VNKSNSKSPK-----EHLEMMKQMSPNKLTHTNFHIPKGGPPAKPGKHSKDKPLKAKG 406
Db 419 IKNELNEKEEELKKNKAHTVEVNTLKEIKL-----LEKKTEDAKEGHKN----- 464
QY 407 RSKGILNGKSTGNSKSPKGLTKPTKM-----KQMTLLDMAKGTQKMTAPRNSG 458
Db 465 -ELNELNQLSKLKE--KDNINKNENTELNDKISSLNSEVNILNKDKQT-----LG 512
QY 459 GTPRTSSPKHKLPPAALHLIAYKENKDREKRSALSCVISKTARLILSSEDRABLPREL 518
Db 513 NDKITLN-----DLINNKNKNEINTSNK-----MKN-----MKEDLAWMLNEEM 550
QY 519 RSL-----VOKRY-----ELLEHKRWASMSSE-----ORKEYLK 548
Db 551 ECKCVVIDEIEKKYKNEIFMLEBEKUKENYADLNDEISILNSIYVKEKEPIEMKEFYE 610
QY 549 KXREELKKLKEKAKERREKEM-----LERLEKQRYEDELGTG 587
Db 611 KNKLNFKNFKSEK-KNIYENELNSRLKYDNEQGLIKQIDELNIQKLKTEBKY--LQLYN 667
QY 588 KULPAFLVDTEGIPNTLFGDVAMVVEFLSCY-----SGLLPDAQPYPTAVLSLMEA 640
Db 668 DNMHFRSICTKIDMPYSENIKSGDLVDFTYAIKRRDESSDANPDPTTHKEMVALEKR 727
QY 641 LSADKGGFLYLRVLVILLQTLLODEIAEDYGLGMKLSIEPLTLHVSSELVRLCL 700
Db 728 HAA-----IVALEEKHEEETAKLEGHEKEVVLRLGE----- 759
QY 701 RRSVDVQEESEGSDTDNDKSAAPEDNEVQDEFLEKLTSEFFELTSEEKLIQILTALCHRI 760
Db 760 ----QHKETIILEEKHKDVVTKLGEQHKENIKLE-----EEHKDVVTKL---- 801
QY 761 LMTYSVDHMETRQQMSAELWTERLAVLKEENDKKRAEKQR-----KEM-EAKNKEN 812
Db 802 -----GDQYKEETAKLEEHAVVVAELEEKHKLGEHKEKEMVDELEKRH 844
QY 813 GKVENGGLKTRKRTVKPEQVDTEAEDMISAVKSRRLAIAQAKKEREIQB----- 864
Db 845 ADFVGL--BEKKAETAKLEEGHKSE-----MNEVEKRHADPVEGLEEKHKAETAKLGECH 899
QY 865 REMVKLERQABE---ERIRKHK---AAAKAFQEBQIAKAKLWMRRTPIGTDNRNRYWL 918
Db 900 REVVAGLEEKHKEVVAELEEKHEETAKLEEGHKEVMAELGEKHEKVAGLEAKEN---- 955
QY 919 FSDVEPGLIEKGWHDSDIDYRPNHKKDHTVSGEDYCPRSKKANLGNQASMTQCH--- 975
Db 956 -----LEEG--HKEMVALEKRRHADLVAVLEEQH-----KAEIHK---LGSEHKEV 996
QY 976 --GTATEVAVETTPKQGNLWFLCDSQKELDELNCLHPOGIRRESQKLERLEKYODII 1033
Db 997 VAGIEKYYKVAIK-----LAEHKDVVTKLGEQHKEEI-----AKLEDGHKEVV 1041
QY 1034 HSIHLARKENLKGCDGNQELLNFI-----RSDLIEVATRLQKGLGYVEETSEFEARYI 1089
Db 1042 NEVE-----KKQASLANLMEENHNKEMIKLEEHKESASDLVEKLYQKDEEVK 1089
```

```
QY 1090 S-----LEKL-----KDFGECVIALQASVIKFKLQGFMAPKQKRRKQSDSASKTERVDSE- 1140
Db 1090 NSNNKIEELTWIKDLSNDSINCYKQILIEV-----EKREYNEINEINKLIVQNEH 1140
QY 1141 ----KKMVEE-----AKVASALEKWKTAIRBAQTFSRHMVLLG-----MLDACIKWDM 1185
Db 1141 KDMNDKILKEKNEIKKLKKNLSNVKVPETKENTYKNSSEMVVNENKERIIVDSVCKENIS 1200
QY 1186 AE-----NARKKVCPPKGGEDDKLILCDECNKAFHFLCLRPALY-----EVPDGE 1229
Db 1201 ESDVEGKGNLKTLSLKKERNIFSIINDNKNESELVDVTIKSAYINKIEMYKKEIED-- 1258
QY 1230 WQCPACQATARRNSRGRNTE-----ESASEDSBDDSDSDEEEEEE----- 1270
Db 1259 -----NGKNIEDLKNKILDSNELINLENMKQNLTDNNNLKKEIEIKDN 1303
QY 1271 --EEEEEEDEYEVAGL 1284
Db 1304 KLNKEKKNENTELNL 1319

RESULT 48
S21801
M:osin heavy chain, neuronal [similarity] - rat
N:Alternate names: myosin II
N:Contains: myosin ATPase (EC 3.6.4.1)
C:Species: Rattus norvegicus (Norway rat)
C:Date: 31-Dec-1993 #sequence revision 31-Dec-1993 #text_change 09-Jul-2004
C:Accession: S21801; P00013; S18134
R:Sun, W.; Chantler, P.D.
J. Mol. Biol. 224, 1185-1193, 1992
A>Title: Cloning of the cDNA encoding a neuronal myosin heavy chain from mammalian brain
A:Reference number: S21801; MUID:92235856; PMID:1569576
A:Accession: S21801
A:Molecule type: mRNA
A:Residues: 1-1999 <SUN>
A:Cross-references: UNIPROT:Q63731; EMBL:X62659
R:Sun, W.; Chantler, P.D.
Biochem. Biophys. Res. Commun. 175, 244-249, 1991
A>Title: A unique cellular myosin II exhibiting differential expression in the cerebral
A:Reference number: P00013; MUID:91151356; PMID:1998509
A:Accession: P00013
A:Molecule type: mRNA
A:Residues: 1914-1998, '1' <SU2>
A:Experimental source: brain
C:Superfamily: myosin heavy chain; myosin motor domain homology
C:Keywords: actin binding; ATP; coiled coil; hydrolase; methylated amino acid; nucleotide
F:84-763/Domain: myosin motor domain homology <MMOT>
F:174-181/Region: nucleotide-binding motif A (P-loop)
F:541-575/Region: actin binding #status predicted
F:653-675/Region: actin binding #status predicted
F:836-1999/Domain: coiled coil #status predicted <COI>
F:836-1276/Region: S2
F:1277-1999/Region: light meromyosin
F:125/Modified site: N6,N6-trimethyllysine (Lys) #status predicted
F:180/Binding site: ATP (Lys) #status predicted
F:693,703/Active site: Cys #status predicted
F:1916/Binding site: phosphate (Ser) (covalent) #status predicted
F:1943/Binding site: phosphate (Ser) (covalent) #status predicted

Query Match 3.2%; Score 259.5; DB 1; Length 1999;
Best Local Similarity 19.3%; Fred. No. 0.00057;
Matches 241; Conservative 230; Mismatches 462; Indels 315; Gaps 56;

QY 75 LLKEEFPAYEKL-VLEMVHNTASLEKLVDTAMLEIMTK-----YAVGECDFGVGKEK 128
Db 793 LARKAFACKQOOLSALKILQRCAYLKLRFHWQWRVTYKVKPLQVTRQEEELQAKDEE 852
QY 129 MLKVKIVK-----IHPLEKVDDEATEKKSDGA----- 155
Db 853 LMK-KVEQTKVEALEEEMERKHQOLLEKKNILAEQLQAEITELFAEAEMRRLAAKKQE 911
QY 156 -----CDSFSSDKENSQIAQDHQKKEVTVVKED--EGRRESINDRARRSPKRLPTSUK 206
```


Db 804 KWAHNDFLVIRNNYWRNSVSIYTVGQECPLYEVPGNSKRNFTTRDFLOVP-IYRL 862
Qy 699 C-----LRRSDVQE-----ESGSDTDNDKSAAFEDNEVQDE-FLEKLETFSEFFEL 744
Db 863 FWKSRDNRPRIRMDIDKAPPAHSESSIRKLKQCADFKTKGMDSNWVWIKPE-----FRL 918
Qy 745 TSEKQLQILTA---LCHRLITWYSVQDHMETQWQSAELWKRERLAVLKEENDKKAEBKQK 801
Db 919 PSEIEIRAWSPQECAYFSM-----IAAQRLKADAGYGEKFLFAPQEDDDEEAQLKL 971
Qy 802 RMEAKAKNGKENGKVLGKTKDRKKRIYKVFEPQVTEAEDMISAVKSRLLAI-----854
Db 972 DDE-----VKVAPWNTR--YIQWGRKCLLOLSGPADPT 1005
Qy 855 -----QAKKEREIOEREMKV-----KLERQABEERIRKHAATAAEKAF 891
Db 1006 GCGEGFSYVRVPNKFTQTKEEQSPKRSVTGTDADLRRLPLQRAKELLRQFKVPEE---1062
Qy 892 QEGIAKAKLVMRRTPIGTDRNHNRYWLFSDDEVPGFLIEKGWVHDSIDYRPNHCKDHTVS 951
Db 1063 -----EIKKLSR-----WEVIDVVRTLSTEKAKAGEBGMKFSRGNRFSIAE 1104
Qy 952 GDEDY---CPR---SKKANLGKNASMNTQHGATVAVETTPPKQGOINLWFLCDSQKELD 1005
Db 1105 HOERYKEBCQRIFDLQNRVLASSEVLSDEAESS-ASESDLEELGKNDENMLSNLSTST 1163
Qy 1006 ELLNCLHPQGITRESOLKREKRYQDIITHSI---HLARKPNILGLKSCDG---NOELLNF 1058
Db 1164 QL-----SREREELER--QELLRLQDDEHGGSPSGGAGKAGKDDPGQOMLAT 1210
Qy 1059 LRSDLIEVATRLQKGLGVETSEFEARVISLEKLKDFGECVIALQASVIKKFLQGWFA 1118
Db 1211 NNQGRILRITRTFRNGD-----KEY-TRVETVRR-QPVIDAIYIKRTTKDQEQFIKQF--1261
Qy 1119 PQKRRKQSDSASAKTEVDEEKKVVEAKVASALEKWKTAJREAQTFSRMHVLGMLDA 1178
Db 1262 -----ATLDEQKEEMKREKRIQEQ--LRRIKRNQERERLQLAQNKQLQFGCHPT 1311
Qy 1179 CIKWMSA-----ENARCKVCPKKGEDDKLILCDECNKAPHFLCLRPALEYVPGGEW 1230
Db 1312 SLGDPKSGSHSKERDSGYKEVPSR-----KKFKL-----KPDCLK 1349
Qy 1231 QCPAC-QPATARNRGRNY-----TEESASESDSEDESEEEEEEEEEEEEDYEVAGLR 1285
Db 1350 KCGACGQVGHMRTNACPLYQMGSLSQSNPSLADDFDEQSEKEMTWDDDLNVNVDGTK 1409
Qy 1286 LRPRKTIRGKSHVIPPAAARSGRRPG---KKPHSTRSQPKAPPVDDAEVDELVLQTKRS 1342
Db 1410 VTLSKILKRHGGDDGKRSGSSGFTLVKPRDA-MGKKRVRGGDLHCDYLQRNKNTAN 1468
Qy 1343 RRQS---LELOKCEBILHKIVKRYFS-WPFREPVTRDEABDYVDYITHPMDPQTQVKNCS 1398
Db 1469 RRRTPVVLVSSILEIHHNLSRMDPVSFPLFPVSAKKVPDYRVVVTKPMDLQTMREYR 1528
Qy 1399 CGSYRSVQBEFLTDMQVFNAAEVYN-CRGSVHLSCHWVTEQCLVLI-----1443
Db 1529 QRRYTSREMFLEDLQIVDNSLIYNGPQSAYTAAQRMFSSCFELLAEREDKMLREKAI 1588
Qy 1444 -----LH-----KHLF-GHPVYRR-KRKKFPDLRAEDGSDPEPAVGOS 1480
Db 1589 NPLDDDDQVALSFIDKLHSQIKQLPESWFLKVPVKNQKQVYTVIKRPMDLLETIGN 1648
Qy 1481 RDEDRRSRAEIQEWLQDTSLSAKINSKDNH 1512
Db 1649 IEAHRVHSRA---EVLADIELIAT--NCEQYN 1675

RESULT 50
148153
myosin heavy chain beta, cardiac muscle [similarity] - golden hamster
C:Species: Mesocricetus auratus (golden hamster)
C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 09-Jul-2004

C:Accession: I48153; A28298
R:Wang, R.; Sole, M.J.; Cukerman, E.; Liew, C.C.
J. Mol. Cell. Cardiol. 26, 1155-1165, 1994
A:Title: Characterization and nucleotide sequence of the cardiac alpha-myosin heavy chain
A:Reference number: I48153; MUID:95115033; PMID:7815459
A:Accession: I48153
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1934 <RES>
A:Cross-references: UNIPROT:P13540; GB:L12104; NID:g402371; PIDN:AAA62313.1; PID:g402372
R:Jandreski, M.A.; Sole, M.J.; Liew, C.C.
Nucleic Acids Res. 16, 4737, 1988
A:Title: Sequence of cDNA encoding the Syrian hamster cardiac beta-myosin heavy chain.
A:Reference number: A28298; MUID:88247788; PMID:3380703
A:Accession: A28298
A:Molecule type: mRNA
A:Residues: 962-965, 'E', 967-980, 'E', 981-985, 'Q', 987-1007, 'A', 1009, 'E', 1011, 'RKT', 1015-1016
536, 'L', 1538-1555, 'K', 1557-1934 <JAN>
A:Cross-references: GB:X07273; NID:g49640; PIDN:CAA30256.1; PID:g49641
A>Note: the authors translated the codon GTG for residue 1504 as Leu
C:Genetics:
A:introns: 66/3; 114/3; 167/1; 176/2; 212/3; 243/3; 265/1; 298/1; 332/3; 379/1; 418/3; 423/3; 1389/2; 1450/3; 1506/1; 1547/3; 1650/3; 1718/3; 1760/3; 1852/3; 1884/3; 1929/3
C:Superfamily: myosin heavy chain; myosin motor domain homology
C:Keywords: actin binding; ATP; cardiac muscle; coiled coil; heart; muscle; nucleotide
F:87-765/Domain: myosin motor domain homology <MOT>
F:177-184/Region: nucleotide-binding motif A (P-loop)
Query Match 3.2%; Score 257.5; DB 2; Length 1934;
Best Local Similarity 18.4%; Pred. No. 0.00067;
Matches 288; Conservative 236; Mismatches 460; Indels 583; Gaps 72;
Qy 249 NKEIVRYFIRNALRAGTGENAPWVVEDELVKYKSLPSKFSDFLLD-----PKY 298
Db 481 NEKLOQFFNH-----MFVLEQEYKKEGIEWTFIDFGMDLQACIDLIEKPMRI 529
Qy 299 MTL-----NPSKRNKNTGSPDRKPSKK-----SKTDNSS 327
Db 530 MSILEEECMFPKATDMTFKALYNHLGKSNFQKPRNVKGQEAHFSLVHYAGTVDNI 589
Qy 328 L-----SSPLNP-----KLWCHVHLKSLGSLPKVKNSKNSKSPPEHLEEM 370
Db 590 LGWLQKNDPLNETVVGLYQKSSLLKLSNLPANVAGADA PVDKGKAKKSGSSFQTVSVL 649
Qy 371 KMSPNKLHTNF-----IPKGGPPAKPKGKSDKPLKAKGRSGKILNGQKSTGN 420
Db 650 HRENKLMTNLRSTPHFVRCIIPNE---TKSPGMDNPLVMHQLRCNGVLEGIRIC--704
Qy 421 SKSPKKGKLT-----PKTKMKQNTLLDMAKGTQKMTRAPNMSGTPTRTSK 466
Db 705 ---RKGFNRLYGDFRQRYRLNPAAIPEGQFIDSRKGAEL-----LSSL 748
Qy 467 PHKLPPAALHUIAYYKEN-----KDREDKRSALSCVISTKARLLSSSDRALPE-ELRS 520
Db 749 DIDHNYKFGTKVFFKAGLLGLLEEMDER--LSRIITR---IQAQSRGLLSRMFPCK 802
Qy 521 LVQKRYELL-----EHKQKWAHSEB--ORKEYL 547
Db 803 LLERDSLVITQWNIAPFMGVKNVPMWMLYFKIKPLKLSAETEKEMATMKEEFGVVKDAL 862
Qy 548 KK---KREELKKK---LKEK-----AKER-----REKEML 571
Db 863 EKSEARRKELEBKWVSLQEKNDLQVQAESQDNADAEERCQDLIKKIKLEAKVKEMT 922
Qy 572 ERLEKQKQYEQELTGKKNLPAPRLVDTPEGLPNTLTFGDVAMVVEFLSCYSGILLPDAQYP 631
Db 923 ERLEDEEEM-NAELTAK-----938
Qy 632 ITAVSLMEALSADKGGFLYNLRVLILQTLLOTLQDEIAB---DYSELGNKLSIPLT 688
Db 939 -----KKRLEDECSLKRDIIDDLTLTAKVEKD 966
Qy 689 LHSVSELVRLCLRRSDVQEESEGSD---TDDNKDSAAAFEDNEVQDEFLEKLTSEFFELT 745

Db 967 KHATENVK-----NUTEEMAGLDEIIIAKLTKEKKALQ--EAHQALDDLQ-----A 1011
Qy 746 SEEKLOILTALCHRIIMTYSQD-----HMETRQOMSAELWKERL-----AVLKEE 791
Db 1012 EEDKVNTLTK--SKVKLEQQVDDLEGSLEQEKVKVRMDLERAKRKLBGDLKLTQESIMDL 1069
Qy 792 NDKKRA-EKOKRMEANENKENVEN--GIGKTDKK-----RIVKFEPOVDTEAEDM 842
Db 1070 NDKQDLDEKLKKDFEL-NALNARIEDEQALQELQKQKLEQALQARIEELEELEAE--R 1125
Qy 843 ISAVKSRRLAIQAKKEREIOER-----EMVKLERQAEERIRK-----HK 884
Db 1126 TARAKVEKLRSDUSRELEISELESELEAGGATSVQIEMNKKREAFQOMRRDLEEA TLQHE 1185
Qy 885 AAA---EKAFQEGIAK-AKLV--MRRTPIGTDRNHNRYMLFSDEVPGLF-----IE 929
Db 1186 ATAALRKKHADSVAEIGEIQDNLQVKQKLEKESEFLELDVDTSNMEQIIKAKANLE 1245
Qy 930 KGWVHDSIDYRFNHCKDHTVSGDEDYCPRSKKANLGNKASNMTOHGTATEVAVETTPK 989
Db 1246 K--MCRTLEDQMNHE-----RSKABETQRSVNDLTSQRAKLQTEG-----1284
Qy 990 QGQNLWFLCDSQKELDE--LLNCLHPQGIRESQ----LKERLE---KRYQDIHSHILA 1039
Db 1285 -----ELSRQDEKEALISQLTGRKLTYYTQOLEDLKRQLEEEVKAKNTLAHALQSA 1335
Qy 1040 RKPNI GLKSCDGNQELLNFRSLDIEVATRLQKGGLYVEETSEFEARVISLEKLDKDFE 1099
Db 1336 R-----HDCD-----LLREQ-----YEEET-EAKAEL-----Q 1357
Qy 1100 CVIALQASVTKFLOQFMAPKQKRKLQSEDSAKTEEVDEBKQV-----EAKVA-- 1150
Db 1358 CVLSKANSEV-----AQWRTKYETDAIQRTTEELEBAKKKLAQRLQDAEEAVEAVN 1407
Qy 1151 ---SALEKWKTAIREAQTSRMMHVLLGMLDACIKWMSAENARCKVCPKKGDDKLILCD 1207
Db 1408 AKCSLSLEKTK-----HRLQNEIED-LMVDVERSNAALADKKQNFDKILAE 1454
Qy 1208 ECKN-----AFHLFCLRPAL-----YEVDPGEWQ 1231
Db 1455 WKQYSESOSELESSQEAESLSLSTELFKLKNAYEESLEHLETFKRENKNLOEISD-----1510
Qy 1232 CPACQPATARNRSGRNYTE-----ESASEDSDESDEE-----1266
Db 1511 -----LTEQLSGTGSIIHELEKIRKQLEBAEKMELOSALEAEASLEHEEGNITLRAQLEF 1564
Qy 1267 ---EEEEEEEEEDYEVAGLRPRKRTINGKHSVIPPAARSGRPPKPKPHSTRSQPKA 1323
Db 1565 NQIKABIERKLAEKDEEMBOAKENHLRVVDSLQTSLDAETRS-----RNEALRV 1613
Qy 1324 PPVDDAEVDELVLQTKRSRRSLOLEQKCEEILHKIVKYRFSWPPFPVTRDEADYDV 1383
Db 1614 KXMEGDLNEMEIQLSHAN-RMAEAQKQVKSLSQSLK-----DTQIQLDDAVRANDD 1665
Qy 1384 ITHPMDFOQVQKSCGYSRSVOEFLTDMKQVFTNAEVYVNCGRSHVLSCHVKTQCCLVVL 1443
Db 1666 LKENIAIVERNNL-----LQAELEELRAVVQEOTERSKLAEQEL---IETSE-RVOL 1714
Qy 1444 LHKHLPCHPYVRKRKPKFPDRLAEDGDSPEAVQSGRDEDRRSRE-----AEIQEW 1495
Db 1715 LHSQ---NTSLINQKKMDADLSQLOTEVE-EAVQECRNAEAKAKAITDAAMMAEELKK 1770
Qy 1496 LODTSLY 1502
Db 1771 EQUTSAH 1777

Search completed: March 9, 2005, 14:25:56
Job time : 90.149 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 9, 2005, 13:18:56 ; Search time 145.036 Seconds
(without alignments)
5405.490 Million cell updates/sec

Title: US-10-702-148-29
Perfect score: 7985
Sequence: 1 MAPLGRKPFPLVNLPGSE.....NCFMMLVNTQFCVALTDTVT 1531

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 65 summaries

Database : Uniprot 03: *
1: uniprot_sprot: *
2: uniprot_trembl: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	7662	96.0	1483	1	BA1B HUMAN	Q9u1g0 homo sapien
2	6977	87.4	1479	1	BA1B MOUSE	Q72277 mus musculus
3	4928	61.7	955	2	Q86UJ6	Q86u16 homo sapien
4	4144	51.9	808	2	Q6PIK4	Q6pi14 homo sapien
5	3728	46.7	1079	2	Q8UVR4	Q8uvr4 xenopus lae
6	3029.5	37.9	657	2	Q8CAU9	Q8cau9 mus musculus
7	2866.5	35.9	777	2	Q6G006	Q6gq06 xenopus lae
8	919.5	11.5	1556	1	BA1A HUMAN	Q9nrl2 h bromodoma
9	668	8.4	1476	2	Q9V9T4	Q9v9t4 drosophila
10	666	8.3	1476	2	Q9NG24	Q9ng24 drosophila
11	665	8.3	1476	2	Q9Y0W1	Q9y0w1 drosophila
12	655	8.2	1357	2	Q7FRP9	Q7frp9 anopheles g
13	654.5	8.2	1427	2	Q9NSL9	Q9nsl9 caenorhabdi
14	640.5	8.0	892	2	Q6P1D9	Q6pid9 mus musculus
15	550	6.9	1698	2	Q6Y194	Q6y194 xenopus lae
16	537	6.7	1972	1	BA2B HUMAN	Q9uif8 homo sapien
17	528.5	6.6	1905	2	Q68D18	Q68d18 homo sapien
18	525.5	6.6	1878	1	BA2A HUMAN	Q9uif5 homo sapien
19	512.5	6.4	1850	1	BA2A MOUSE	Q9uif5 mus musculus
20	507	6.3	718	2	Q8C8D1	Q8cd81 mus musculus
21	503	6.3	1202	2	Q7QBK2	Q7qbk2 anopheles g
22	503	6.3	2130	1	BA2B CHICK	Q9del3 gallus gall
23	487	6.1	2060	2	Q7Q3S9	Q7q3s9 anopheles g
24	463.5	5.8	1370	2	O45075	O45075 caenorhabdi
25	449.5	5.6	1376	2	Q23590	Q23590 caenorhabdi
26	407	5.1	627	1	BA1A XENLA	Q8uvr5 xenopus lae
27	371.5	4.7	1508	2	Q75JF5	Q75jpf5 dictyosteli
28	361.5	4.5	10578	2	Q8ISF5	Q8isf5 caenorhabdi
29	361.5	4.5	18519	2	Q8ISF6	Q8isf6 caenorhabdi
30	361.5	4.5	18534	2	Q8ISF7	Q8isf7 caenorhabdi
31	353	4.4	886	2	Q80U42	Q80u42 mus musculus

32	352.5	4.4	7210	2	Q9V7G8	Q9v7g8 drosophila
33	352.5	4.4	9270	2	Q8MLD9	Q8mld9 drosophila
34	345.5	4.3	1364	1	ITC1 YEAST	P53125 saccharomyc
35	342	4.3	2759	2	Q9VID9	Q9vid9 drosophila
36	342	4.3	3080	2	Q9V602	Q9v602 drosophila
37	335.5	4.2	1637	2	Q75PR8	Q75pr8 hemiceutrot
38	331.5	4.2	3109	2	Q9BNQ0	Q9bnq0 drosophila
39	329	4.1	617	2	Q6Z178	Q6z178 oryza sativ
40	329	4.1	1230	2	Q6BV90	Q6bv90 debaryomyce
41	329	4.0	1790	1	USO1 YEAST	P25386 saccharomyc
42	322	4.0	1268	2	Q7LGT1	Q7lgt1 saccharomyc
43	322	4.0	1790	2	Q07380	Q07380 saccharomyc
44	321.5	4.0	969	2	Q6CAZ9	Q6caz9 yarrowia li
45	319.5	4.0	1980	2	Q6FWE0	Q6fwe0 candida gla
46	317.5	4.0	2042	2	Q6BUQ9	Q6bug9 debaryomyce
47	316	4.0	1596	2	Q8IJ44	Q8ij44 plasmodium
48	314	3.9	1661	2	Q06166	Q06166 plasmodium
49	310.5	3.9	1510	2	Q25920	Q25920 plasmodium
50	310.5	3.9	1864	2	Q6P5H2	Q6psh2 mus musculus
51	307	3.8	3616	2	Q9W6V0	Q9w6v0 gallus gall
52	302.5	3.8	1345	2	Q9Y2L2	Q9y2l2 homo sapien
53	302.5	3.8	1972	2	Q8R384	Q8r384 mus musculus
54	302.5	3.8	1984	2	Q69ZX3	Q69zx3 mus musculus
55	302.5	3.8	3259	1	GOB1 HUMAN	Q14789 homo sapien
56	301.5	3.8	683	2	Q960I3	Q960i3 drosophila
57	301	3.8	2492	1	ATRX PONPY	Q7ygm3 pongo pygma
58	299.5	3.8	1431	2	Q9GT23	Q9gt23 homo sapien
59	299	3.7	2492	1	ATRX HUMAN	P46100 homo sapien
60	299	3.7	3111	2	Q9VH10	Q9vh10 drosophila
61	298.5	3.7	1972	1	MYHB MOUSE	Q08638 mus musculus
62	297.5	3.7	1927	2	Q75WX9	Q75wx9 helicobacte
63	296.5	3.7	2116	1	MYSD D1CDI	P08799 dictyosteli
64	295.5	3.7	400	2	Q95RV8	Q95rv8 drosophila
65	295.5	3.7	2081	2	Q9LH98	Q9lh98 arabidopsis

ALIGNMENTS

RESULT 1	BA1B HUMAN	STANDARD;	PRT; 1483 AA.
ID	Q9U1G0; Q95039; Q95247; Q95277;		
AC	28-FEB-2003 (Rel. 41, Created)		
DT	28-FEB-2003 (Rel. 41, Last sequence update)		
DT	25-OCT-2004 (Rel. 45, Last annotation update)		
DE	Bromodomain adjacent to zinc finger domain protein 1B (Williams-Beuren syndrome chromosome region 9 protein) (WBSCR9) (Williams syndrome transcription factor) (hWALP2).		
GN	Name=BAZ1B; Synonyms=WBSCR10, WBSCR9, WSTF;		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
NCBI_TaxID=9606;			
[1]	SEQUENCE FROM N.A. (ISOFORM 1).		
RX	MEDLINE=99077764; PubMed=9858827;		
RA	Peoples R.J., Cisco M.J., Kaplan P., Francke U.;		
RT	"Identification of the WBSCR9 gene, encoding a novel transcriptional regulator, in the Williams-Beuren syndrome deletion at 7q11.23.";		
RL	Cytogenet. Cell Genet. 82:238-246(1998).		
RN	[2]		
RP	SEQUENCE FROM N.A. (ISOFORM 1).		
RX	MEDLINE=99047530; PubMed=9828126; DOI=10.1006/geno.1998.5578;		
RA	Lu X., Meng X., Morris C.A., Keating M.T.;		
RT	"A novel human gene, WSTF, is deleted in Williams Syndrome.";		
RL	Genomics 54:241-249(1998).		
RN	[3]		
RP	SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).		
RC	TISSUE=Testis;		
RX	MEDLINE=20130112; PubMed=10662543; DOI=10.1006/geno.1999.6071;		
RA	Jones M.H., Hamana N., Nezu J., Shimane M.;		
RT	"A novel family of bromodomain genes.";		

```

RL Genomics 63:40-45(2000).
RN [4]
RP SEQUENCE FROM N.A.
RA Minx P., Graves T., Duckels G., Harrison M., Waterston R.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
RN [5]
RP FUNCTION.
RX MEDLINE=21977304; PubMed=11980720; DOI=10.1093/emboj/21.9.2231;
RA Bozenok L., Wade P.A., Varga-Weisz P.;
RL "WSF-ISWI chromatin remodeling complex targets heterochromatic
RT replication foci.";
RL EMBO J. 21:2231-2241(2002).
CC -!- FUNCTION: Forms a chromatin remodeling complex that mobilizes
CC nucleosomes and reconfigures irregular chromatin to a regular
CC nucleosomal array structure.
CC -!- SUBUNIT: Interacts with ISWI (imitation SWI protein) to form the
CC WSTF-ISWI chromatin remodeling complex (WICH).
CC -!- SUBCELLULAR LOCATION: Nuclear (Potential). Accumulates in
CC pericentromeric heterochromatin during replication.
CC -!- ALTERNATIVE PRODUCTS:
CC Event-Alternative splicing; Named isoforms=2;
CC Name=1;
CC IsoId=Q9UIG0-1; Sequence=Displayed;
CC Name=2;
CC IsoId=Q9UIG0-2; Sequence=VSP 000552;
CC -!- TISSUE SPECIFICITY: Ubiquitously expressed with high levels of
CC expression in heart, brain, placenta, skeletal muscle and ovary.
CC -!- DEVELOPMENTAL STAGE: Expressed at equal levels in 19-23 weeks old
CC fetal tissues.
CC -!- DISEASE: Haploinsufficiency of BAZ1B may be the cause of certain
CC cardiovascular and musculo-skeletal abnormalities observed in
CC Williams-Beuren syndrome (WBS), a rare developmental disorder. It
CC is a contiguous gene deletion syndrome involving genes from
CC chromosome band 7q11.23.
CC -!- SIMILARITY: Belongs to the WAL family.
CC -!- SIMILARITY: Contains 1 bromodomain.
CC -!- SIMILARITY: Contains 1 DDT domain.
CC -!- SIMILARITY: Contains 1 PHD-type zinc finger.
CC -!- SIMILARITY: Contains 1 PHD-type zinc finger.
CC -!- SIMILARITY: Contains 1 WAC domain.
CC -!- CAUTION: Ref.2 sequence differs from that shown due to frameshifts
CC in positions 1031, 1042 and 1422.
CC -!- CAUTION: Ref.3 sequence differs from that shown due to a
CC frameshift in position 1478.
CC -!- CAUTION: Ref.4 sequence differs from that shown due to erroneous
CC gene model prediction.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF084479; AAC08675.1; -;
DR EMBL; AF072810; AAC97879.1; ALT_FRAME.
DR EMBL; AF032253; BAA89210.1; ALT_FRAME.
DR EMBL; AC005074; AAD04720.1; ALT_SEQ.
DR EMBL; AC005089; -; NOT_ANNOTATED_CDS.
DR PDB; 1F62; NMR; A=1185-1235.
DR TRANSFAC; T04145; -;
DR Genew; HGNC:961; BAZ1B.
DR MIM; 605681; -;
DR GO; GO:0005634; C:nucleus; NAS.
DR GO; GO:0003700; F:transcription factor activity; NAS.
DR GO; GO:0008270; F:zinc ion binding; NAS.
DR GO; GO:0006350; P:transcription; NAS.
DR InterPro; IPR001487; Bromodomain.
DR InterPro; IPR004022; DDT.
DR Pfam; PF00439; Bromodomain; 1.
DR Pfam; PF00628; PHD; 1.
DR PRINTS; PR00503; BROMODOMAIN.

```

```

DR SMART; SM00297; BROMO; 1.
DR SMART; SM00571; DDT; 1.
DR SMART; SM00249; PHD; 1.
DR PROSITE; PS00633; BROMODOMAIN_1; FALSE_NEG.
DR PROSITE; PS00014; BROMODOMAIN_2; 1.
DR PROSITE; PS50827; DDT; 1.
DR PROSITE; PS01359; ZF_PHD_1; 1.
DR PROSITE; PS50016; ZF_PHD_2; 1.
KW 3D-structure; Alternative splicing; Bromodomain; Coiled coil;
KW Nuclear protein; Transcription regulation; Williams-Beuren syndrome;
KW Zinc-finger.
FT DOMAIN 20 126 WAC.
FT DOMAIN 604 668 DDT.
FT ZN_FING 1184 1234 PHD-type.
FT DOMAIN 1356 1426 Bromodomain.
FT DOMAIN 1356 578 Lys-rich.
FT DOMAIN 533 586 Coiled coil (Potential).
FT DOMAIN 768 814 Coiled coil (Potential).
FT DOMAIN 850 893 Coiled coil (Potential).
FT DOMAIN 1245 1283 Coiled coil (Potential).
FT DOMAIN 1261 1273 Poly-Glu.
FT VARSPLIC 660 663 Missing (in isoform 2).
FT CONFLICT 14 14 K -> N (in Ref. 3).
FT CONFLICT 22 22 L -> F (in Ref. 3).
FT CONFLICT 136 136 K -> E (in Ref. 1).
FT CONFLICT 191 191 N -> D (in Ref. 4).
FT CONFLICT 298 298 Y -> V (in Ref. 4).
FT CONFLICT 823 823 E -> R (in Ref. 3).
FT CONFLICT 1191 1191 R -> P (in Ref. 3).
FT CONFLICT 1354 1354 K -> M (in Ref. 2).
FT CONFLICT 1438 1438 A -> V (in Ref. 3).
SQ SEQUENCE 1483 AA; 170902 MW; 0CCI46FEBB954261 CRC64;

Query Match 96.0%; Score 7662; DB 1; Length 1483;
Best Local Similarity 99.0%; Pred. No. 3.1e-307;
Matches 1474; Conservative 2; Mismatches 7; Indels 6; Gaps 2;

QY 1 MAPLGRKPPPLVNPDPGEEPPFTTPTQEAFTREEYEARLERYSERIWTCKSGSSQL 60
DB 1 MAPLGRKPPPLVNPDPGEEPLFTTPTQEAFTREEYEARLERYSERIWTCKSGSSQL 60
QY 61 THKGAWEEOEVAELLKEEPPAWYKLVLEWVHNTASLEKLVDTAWLEIMTKYAVGSEC 120
DB 61 THKGAWEEOEVAELLKEEPPAWYKLVLEWVHNTASLEKLVDTAWLEIMTKYAVGSEC 120
QY 121 DFEVGKRMVKVIVKIHPLEKVDDEATEKSDGACDSPSSDKENSSQIAQDHQKKEVTV 180
DB 121 DFEVGKRMVKVIVKIHPLEKVDDEATEKSDGACDSPSSDKENSSQIAQDHQKKEVTV 180
QY 181 KEDEGRRESINDRARRSPRKLPSTLKKGERKWAQPKPLPHKYDVVKLQNEDKLISNVPADS 240
DB 181 KEDEGRRESINDRARRSPRKLPSTLKKGERKWAQPKPLPHKYDVVKLQNEDKLISNVPADS 240
QY 241 LIRTERPPNKEIVRYFIRHNALRAGTGNAPWVEDELVKVYSLPSKPSDFLLDPKYWT 300
DB 241 LIRTERPPNKEIVRYFIRHNALRAGTGNAPWVEDELVKVYSLPSKPSDFLLDPKYWT 300
QY 301 LNPSTKRKNTGSPDRKPSKSKTNDSSLSPLNPKLWCHVHLKKSLSGSPKVKNSKNSK 360
DB 301 LNPSTKRKNTGSPDRKPSKSKTNDSSLSPLNPKLWCHVHLKKSLSGSPKVKNSKNSK 360
QY 361 SPEEHLEEMOMSPNKLHTNFHIPKGGPPAKKPKGKSDKPLKAGRSKGIINGOKSTGN 420
DB 361 SPEEHLEEMOMSPNKLHTNFHIPKGGPPAKKPKGKSDKPLKAGRSKGIINGOKSTGN 420
QY 421 SKSPKKGKLTPTKTKMQLTLDMAKGTOKMTRAPNSGTTPTSTSSKPHKHLPPAALHLIA 480
DB 421 SKSPKKGKLTPTKTKMQLTLDMAKGTOKMTRAPNSGTTPTSTSSKPHKHLPPAALHLIA 480
QY 481 YYKENKREDKRSALSCVISKTARLLSSEDRLRPEELRSLVQKRYELLEHKRWASMS 540
DB 481 YYKENKREDKRSALSCVISKTARLLSSEDRLRPEELRSLVQKRYELLEHKRWASMS 540

```



```
QY 541 EQRKEVYKKRBEELKKKLEKAKERREKEMLEKRYEDQELTGKMLPAPRLVDTPE 600
DB 541 EQRKEVYKKRBEELKKKLEKAKERREKEMLEKRYEDQELTGKMLPAPRLVDTPE 600
QY 601 GLPNTLFGDVAMVVEFLSCVSGILLDPADQVPIITAVSLMEALSADKGGFLYLNRLVILQ 660
DB 601 GLPNTLFGDVAMVVEFLSCVSGILLDPADQVPIITAVSLMEALSADKGGFLYLNRLVILQ 660
QY 661 TLLQTLQDIEADYELGELGMLKSEIPLTLHSSELVRLCLRRSDVQEESEGSDTDNKS 720
DB 658 -LLQTLQDIEADYELGELGMLKSEIPLTLHSSELVRLCLRRSDVQEESEGSDTDNKS 716
QY 721 AAPEDNEVDFLEKLETSSEFFELTSEKQLIQTALCHRLMTYSVQDHNMTQQMSAEL 780
DB 717 AAPEDNEVDFLEKLETSSEFFELTSEKQLIQTALCHRLMTYSVQDHNMTQQMSAEL 776
QY 781 WKERLAVLKEENDKKAEKQKREMEAKNKENGKVGKTKRKRIVKFPQVDTEAE 840
DB 777 WKERLAVLKEENDKKAEKQKREMEAKNKENGKVGKTKRKRIVKFPQVDTEAE 836
QY 841 DMISAVKSRLLAIAQAKEREIQEREMKVKLERQAEERIRKHAAAEKAFQEGIAKAL 900
DB 837 DMISAVKSRLLAIAQAKEREIQEREMKVKLERQAEERIRKHAAAEKAFQEGIAKAL 896
QY 901 VMRTPIGTRNHRNRYLFSDEVPGLFIEKGWVHDSIDYRFNHHCKDHTVSGDEDYCPRS 960
DB 897 VMRTPIGTRNHRNRYLFSDEVPGLFIEKGWVHDSIDYRFNHHCKDHTVSGDEDYCPRS 956
QY 961 KXANLGKNSMNTQHGTAETAVETTPKQGNLWFLCDSQKELDBLNLCHLPQGIRESQ 1020
DB 957 KXANLGKNSMNTQHGTAETAVETTPKQGNLWFLCDSQKELDBLNLCHLPQGIRESQ 1016
QY 1021 LKERLEKRYQDIITHSLHARKNPLGLKSCDQGNLFLNLSDLIEVATRLQKGLGYVEE 1080
DB 1017 LKERLEKRYQDIITHSLHARKNPLGLKSCDQGNLFLNLSDLIEVATRLQKGLGYVEE 1076
QY 1081 TSEFEARVLSLEKLDGFCGVIALQASVIKKFLQGFMAPKQKRKLQSDSARTEEVDEE 1140
DB 1077 TSEFEARVLSLEKLDGFCGVIALQASVIKKFLQGFMAPKQKRKLQSDSARTEEVDEE 1136
QY 1141 KKWVEAKVASALEKWKTAIREAQFTSRMHVLLGMLDACIKWMSAENARCKVCPKGGD 1200
DB 1137 KKWVEAKVASALEKWKTAIREAQFTSRMHVLLGMLDACIKWMSAENARCKVCPKGGD 1196
QY 1201 DXLLILDCENKAFHLCFLRPALYVDPGEWQCPACOPATARNRSGRNTYESASDSD 1260
DB 1197 DXLLILDCENKAFHLCFLRPALYVDPGEWQCPACOPATARNRSGRNTYESASDSD 1256
QY 1261 DESDEEEEEEEDYEVAGLRRLRPRTIRGKHSVIPPAAARSRRFPKPKPHSTRSQ 1320
DB 1257 DESDEEEEEEEDYEVAGLRRLRPRTIRGKHSVIPPAAARSRRFPKPKPHSTRSQ 1316
QY 1321 KPAPPVDDAEVDDELVLQTRSSRSQSLQKEEILHKIVKTRFWPPFPVTRDEADY 1380
DB 1317 KPAPPVDDAEVDDELVLQTRSSRSQSLQKEEILHKIVKTRFWPPFPVTRDEADY 1376
QY 1381 YDVITHPMDFTQVQNKSCGYSRVSQERLTDMKQVFTNAEVNCRGSHVLSQWVTEOCL 1440
DB 1377 YDVITHPMDFTQVQNKSCGYSRVSQERLTDMKQVFTNAEVNCRGSHVLSQWVTEOCL 1436
QY 1441 VYLLHKLHFGHPYVRKRKKKFPDLRAEDGSDPEAVGQSRDEDRSRE 1489
DB 1437 VALLHKLHFGHPYVRKRKKKFPDLRAEDGSDPEAVGQSR--GRQKX 1483

RESULT 2
ID BA1B MOUSE
AC Q92377: Q9CU68; STANDARD; PRT; 1479 AA.
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
```

Bromodomain adjacent to zinc finger domain protein 1B (Williams-Beuren syndrome chromosome region 9 protein homolog) (WBS9).

Name=Baz1b; Synonyms=Wbscr9;
Mus musculus (Mouse).
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxID=10090;
[1]
SEQUENCE FROM N.A.
RP MEDLINE=99077764; PubMed=9858827;
RA Peoples R.J., Cisco M.J., Kaplan P., Francke U.;
RT Identification of the WBS9 gene, encoding a novel transcriptional
regulator, in the Williams-Beuren syndrome deletion at 7q11.23;
RL Cytogenet. Cell Genet. 82:238-246(1998).
RN [2]
SEQUENCE OF 1-482 FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Thymus;
RX MEDLINE=22354683; PubMed=12466851; DOI=10.1038/nature01266;
RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
Nikaido I., Otsu N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,
Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,
Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,
Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,
Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,
Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,
Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,
Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,
Kongaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,
Nagashima T., Nunata K., Okido T., Pavan W.J., Pertea G., Pesole G.,
Petrowsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,
Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,
Sandelin A., Schneider C., Sempile C.A., Setou M., Shinada K.,
Sulana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,
Verardo R., Wegner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,
Wilning L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,
Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,
Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,
Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,
Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,
Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,
Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
Birney E., Hayashizaki Y.;
RT "Analysis of the mouse transcriptome based on functional annotation of
60,770 full-length cDNAs";
RL Nature 420:563-573(2002).
CC -!- FUNCTION: Forms a chromatin remodeling complex that mobilizes
nucleosomes and reconfigures irregular chromatin to a regular
nucleosomal array structure (By similarity).
CC -!- SUBUNIT: Interacts with ISWI (imitation SWI protein) to form the
WSTF-ISWI chromatin remodeling complex (WICH) (By similarity).
CC -!- SUBCELLULAR LOCATION: Nuclear (Potential). Accumulates in
pericentromeric heterochromatin during replication (By
similarity).
CC -!- TISSUE SPECIFICITY: Expressed in all tissues examined including
heart, brain, spleen, lung, skeletal muscle, kidney and testis.
CC -!- DEVELOPMENTAL STAGE: Expressed as early as day 7 and in equal
amounts during gestation.
CC -!- SIMILARITY: Belongs to the WAF family.
CC -!- SIMILARITY: Contains 1 bromodomain.
CC -!- SIMILARITY: Contains 1 DDT domain.
CC -!- SIMILARITY: Contains 1 PHD-type zinc finger.
CC -!- SIMILARITY: Contains 1 WAC domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
or send an email to license@isb-sib.ch).
CC -----

DR EMBL; AF084480; AAD08676.1; -.
DR EMBL; AK017894; BAB30992.1; -.
DR PIR; T17401; T17401.
DR HSP; Q9UIG0; IF62.
DR MGD; MG1:J353499; Baz1b.
DR GO; GO:0005721; C:centric heterochromatin; IDA.
DR GO; GO:0000793; C:condensed chromosome; IDA.
DR GO; GO:0005634; C:nucleus; IDA.
DR GO; GO:0005515; F:protein binding; IPI.
DR GO; GO:0006333; P:chromatin assembly/disassembly; IDA.
DR GO; GO:0006338; P:chromatin remodeling; IDA.
DR InterPro; IPR001487; Bromodomain.
DR InterPro; IPR004022; DDT.
DR InterPro; IPR001965; Znf_PHD.
DR Pfam; PF00628; PHD; 1.
DR Pfam; PF00439; Bromodomain; 1.
DR PRINTS; PR00503; BROMODOMAIN.
DR SMART; SM00297; BROMO; 1.
DR SMART; SM00571; DDT; 1.
DR SMART; SM00249; PHD; 1.
DR PROSITE; PS00633; BROMODOMAIN_1; FALSE_NEG.
DR PROSITE; PS50014; BROMODOMAIN_2; 1.
DR PROSITE; PS50827; DDT; 1.
DR PROSITE; PS01359; ZF_PHD_1; 1.
DR PROSITE; PS50016; ZF_PHD_2; 1.
KW Bromodomain; Coiled Coil; Nuclear protein; Transcription regulation;
KW Zinc-finger.
FT DOMAIN 20 126 WAC.
FT DOMAIN 605 669 DDT.
FT ZN_FING 1184 1234 PHD-type.
FT DOMAIN 1352 1422 Bromodomain.
FT DOMAIN 537 587 Coiled coil (potential).
FT DOMAIN 769 815 Coiled coil (potential).
FT DOMAIN 851 895 Coiled coil (potential).
FT DOMAIN 1257 1284 Coiled coil (potential).
SQ SEQUENCE 1479 AA; 170788 MW; DFB5816AE1C0634F CRC64;

Query Match 87.4%; Score 6977; DB 1; Length 1479;
Best Local Similarity 90.0%; Pred. No. 4.9e-279;
Matches 1343; Conservative 67; Mismatches 66; Indels 16; Gaps 8;

QY 1 MAPLGRKFPFLVNPDPGEEPPFTTPTQEAFTREEYEARELRYSERIWTCKSGSSQL 60
Db 1 MAPLGRKFPFLVNPDPGEEPLFTTPTQEAFTREEYEARELRYSERIWTCKSGSSQL 60
QY 61 THKEAWEEQEAELLKEEFPNWEKLVLEMVHHTASLEKLVDTAWLEIMTKYAVGSEC 120
Db 61 THKEAWEEQEAELLKEEFPNWEKLVLEMVHHTASLEKLVDSAWLEIMTKYAVGSEC 120
QY 121 DPEVKEKMKLVKVIKIHPLEKVDDEATEKSDGACDSSDKENSSOIAODHOKETVV 180
Db 121 DPEVKEKMKLVKVIKIHPLEKVDDEAVEKSDGACDSSDKENSSQMAODLQKKTVV 180
QY 181 KEDEGRRESINDRARRSRKLTSLKGERKWAPKFLPHKYDVKLQNEDKIISNPADS 240
Db 181 KEDEGRRESINDRARRSRKLTSLKGERKWAPKFLPHKYDVKLQNEDKIISNPADS 240
QY 241 LIRTERPNKEIVRFIRHNALRAGTGENAPWVEDELVKYSPSPKFSDFLLDPKYMT 300
Db 241 LIRTERPNKEILRVFIRHNALRAGTGENAPWVEDELVKYSPSPKFSDFLLDPKYMT 300
QY 301 LNPSTKRNTGSDPRKPSKSTONSLSPLNPKLWCHVHLKSLGSPKLVKNSKNSK 360
Db 301 LNPSTKRNTGSDPRKPSKPKRDSLSPLNPKLWCHVHLKSLGSPKLVKNSKNSK 360
QY 361 SPEEHLEEMQKMSF--NKLHTNFHPPKGGPAKKPGKSHSKPLKAKGRSGKILNGQKST 418
Db 361 SPEEHLEGVKIMSPNNKLH-SFHIPKGPAAKPGKSHSKPLKAKGRSGKILNGQKST 419
QY 419 GNSKSPKGLTPPTKTKMQLTLLDMAKGTQKMTAPRNSGGTPTRTSSKPKHLLPPAALHL 478
Db 420 GNSKSPSKVTKPTKTKMQLTLLDMAKGTQKMTAPRNSGGTPTRTSSGKPKHLLPPAALHL 479

QY 479 IAYYKKNKDRDKESALSVCISKTRALLSSSEDRARLPPELRSLVQKRYELLEHKRWASW 538
Db 480 IAYYKKNKDKEDKSALSVCISKTRALLSSNEDRALPELRALVQKRYELLEHKRWASW 539
QY 539 SEEQKEYLKKRBEELKKLKEKAKERREKEMLERLEKQKRYEQELTGKQLPAFLVDT 598
Db 540 SEEQKEYLKKRQELKERLEKAKERREKEMLERLEKQKRYEQELTGKQLPAFLVDT 599
QY 599 PEGLPNTLFGDVAMVVEFLSCYSGLLLPDAQYPTTAVSLMEALSADKGGFLYLNRVLVL 658
Db 600 PEGLPNTLFGDVALVVEFLSCYSGLLLPDAQYPTTAVSLMEALSADKGGFLYLNRVLVI- 658
QY 659 LQTLLOTLQDBIAEDYGEIMKMLSEPLTLHSHSELVRLCLRRSDVDEEGSDTDDNK 718
Db 659 ---LLQTLLOQDEIAEDYGEIMKMLSEPLTLHSHSELVRLCLRRCDVDEEGSDTDDNK 715
QY 719 DSAAFEDNEVDQEFLEKLETSEFFELTSEEKLOILTALCHRILMTYSVQDHMETRQMSA 778
Db 716 DSTPFEDNEVDQEFLEKLETSEFFELTSEEKLRILTALCHRILMTYSVQDHMETRQVSA 775
QY 779 ELWKERLAVLKEENDKKAEEKQKKEAKKENGKVENGLKTDKRIYKVFEPQVDT 838
Db 776 ELWKERLAVLKEENDKKAEEKQKKEAKKENGKVENGLKTDKRIYKVFEPQVDT 835
QY 839 AEDMISAVKSRLLAIQAKKEREIQEREMKVKLEKROAEEERI RKHAAAEEKAFQBGIAKA 898
Db 836 ADDMISAVKSRLLSMQAKRKREIQERETKVKLEKREAEERERKHAAAEEKAFQBGIAKA 895
QY 899 KLVNRRTPIGTDNRNRYWLFSDVPGLFIEKGWVHDSIDYRFNHCHKXDTVSGDEYCP 958
Db 896 KLVNRRTPIGTDNRNRYWLFSDVPGLFIEKGWVHDSIDYRFKHHRKXDNLPDDYCP 955
QY 959 RSKKANLGKNASMTQHGTA TEVAVETTPKQGNLWFLCDSQKELDLNCLHPQGTRE 1018
Db 956 RRKKANLGKNASVNAHGPAL- AVETTVPRQGNLWFLCDSQKELDLNCLHPQGTRE 1014
QY 1019 SOLKLEKRYQDIITHSLARKPNLGIKSCDGNQELLNFI RSLIEVATLQKGLGVV 1078
Db 1015 SQFERLEKRYQEIITHCNMYMARKPNLGIKSCDGNQELLNFI RSLIEVATLQKGLGYM 1074
QY 1079 EETSEFEARVISLEKLDGFCVIALQASVIKKFLQGFMAKPKQKRKLQSDS AKTEEVD 1138
Db 1075 EGTSEFEARVISLEKLDGFCVIALQASVIKKFLQGFMAKPKQKRKLQSDS AKTEEVD 1134
QY 1139 BEKQWBEAKVASALEKWKTAI REAQTFPSRHHVLLGMLDACIKWDMSENARCKVCPKG 1198
Db 1135 BEKQWBEAKVASALEKWKTAI REAQTFPSRHHVLLGMLDACIKWDMSENARCKVCKKG 1194
QY 1199 EDDKLLCDECNKAPHLFCLRPALYVDPDGEWQCPACOPATARNRSRGRNTYESASBDS 1258
Db 1195 EDDKLLCDECNKAPHLFCLRPALYVDPDGEWQCPACOPATARNRSRGRNTYESBDS 1254
QY 1259 EDDDS--DEEEEBEEEEEDYEVAGLRPRKTIRGKHSVI PPAARSGRPGKKPHSTR 1317
Db 1255 EGDESGEEEEEEEEEDYEVAGLRPRKTIRGKHSVI - PAARSGRPGKKSHPAR 1313
QY 1318 RSQKAPVDDAEVDDELVLQTKRSRRQSLEOKCEEILHKLVKRYRSPREPVTREDA 1377
Db 1314 RSRPK---DDPEVDDLVLQTKRISRRQSLEQCEDELHKLVKRYRSPREPVTREDA 1369
QY 1378 EDYDVDTHTPMDFTQVQNKSCGSYRSVQBEFLTDMKOVFTNAEVVNCRSHVLSQWVTE 1437
Db 1370 EDYDVDTHTPMDFTQVQNKSCGSYRSVQBEFLTDMKOVFTNAEVVNCRSHVLSQWVTE 1429
QY 1438 QCLVLLVHLKHLPGHPYVRRKKEKFPDRLAEDGDSSEPAVGQSRDEDRSR 1489
Db 1430 QCLLALLQKHLPGHPYVRRKKEKFPDRLAEDGDSSESVQSR--GRQKK 1479

RESULT 3
Q86UJ6
ID Q86UJ6 PRELIMINARY; PRT; 955 AA.
AC Q86UJ6;

DT 01-JUN-2003 (TrEMBLrel. 24, Created)
 DT 01-JUN-2004 (TrEMBLrel. 24, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE Hypothetical protein BAZ1B (Fragment).
 GN Name=BAZ1B;
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99063792; PubMed=9847074;
 RA Wilson R.;
 RT "Toward a complete human genome sequence.";
 RL Genome Res. 8:1097-1108(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Graves T., Hawkins M., Maas J.;
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Waterston R.;
 RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Waterston R.H.;
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE FROM N.A.
 RA Waterston R.;
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
 RN [6]
 RP SEQUENCE FROM N.A.
 RA Waterston R.;
 RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AC005089; AAP22332.1; -.
 DR InterPro; IPR004022; DDT.
 DR SMART; SM00571; DDT; 1.
 DR PROSITE; PS0827; DDT; 1.
 KW Hypothetical protein.
 FT NON TER 955 955
 SQ SEQUENCE 955 AA; 110259 MW; 872A54E554424002 CRC64;

Query Match 61.7%; Score 4928; DB 2; Length 955;
 Best Local Similarity 99.3%; Pred. NO. 6.2e-195;
 Matches 952; Conservative 0; Mismatches 3; Indels 4; Gaps 1;

QY 1 MAPLGRKDFPLVNPPLGPEPPFTIPIHTQEAFTREYEAEARLERYSERIWTCKSTGSSQL 60
 DB 1 MAPLGRKDFPLVNPPLGPEPPFTIPIHTQEAFTREYEAEARLERYSERIWTCKSTGSSQL 60
 QY 61 THKEAWEERQEVALLKEBFPWYKLVLEMHVHNTASLEKLVDTAWLEIMTKYAVGERC 120
 DB 61 THKEAWEERQEVALLKEBFPWYKLVLEMHVHNTASLEKLVDTAWLEIMTKYAVGERC 120
 QY 121 DFEVGRKMLKVKIKVHPLEKVEDATEKSDGACDSSDKENSSQIAQHQKKEVTV 180
 DB 121 DFEVGRKMLKVKIKVHPLEKVEDATEKSDGACDSSDKENSSQIAQHQKKEVTV 180
 QY 181 KEDEGRRESINDARRSPKPLPTSLKGRKWAAPPKFLPHKYDVVKLQNEDKIISNPADS 240
 DB 181 KEDEGRRESINDARRSPKPLPTSLKGRKWAAPPKFLPHKYDVVKLQNEDKIISNPADS 240
 QY 241 LIRTERPPNKEIVRYFIRNALRAGTGGENAPVVEDELVKYSLPSKFSDFLLDPYKNT 300
 DB 241 LIRTERPPNKEIVRYFIRNALRAGTGGENAPVVEDELVKYSLPSKFSDFLLDPYKNT 300
 QY 301 LNPSTKRNTGSPDRKPSKSKTDNSLSPLNPKLWCHVHLKXLSGSPKLVKXSKNSK 360
 DB 301 LNPSTKRNTGSPDRKPSKSKTDNSLSPLNPKLWCHVHLKXLSGSPKLVKXSKNSK 360
 QY 361 SPEEHLLEEMKMMSPNKLHTNFHPKGGPPAKPGKHSDKPLKAGRSKGILNGQKSTGN 420

Db 361 SPEEHLLEEMKMMSPNKLHTNFHPKGGPPAKPGKHSDKPLKAGRSKGILNGQKSTGN 420
 QY 421 SKSPKKGLKTPKTQKMQMTLLDMAKGTQKMTAPRNSGGTPTRTSSPKHKLPLPAALHLIA 480
 Db 421 SKSPKKGLKTPKTQKMQMTLLDMAKGTQKMTAPRNSGGTPTRTSSPKHKLPLPAALHLIA 480
 QY 481 YKKNKDRDKRSALSVCVISTARLLSSDDRARLPPELRLSLVQRYELLEHHKRWASMS 540
 Db 481 YKKNKDRDKRSALSVCVISTARLLSSDDRARLPPELRLSLVQRYELLEHHKRWASMS 540
 QY 541 EORKEYLKKRBEELKKLKKAKEREKEMLEKQKRYEDQELTGKNLPARFLVDTP 600
 Db 541 EORKEYLKKRBEELKKLKKAKEREKEMLEKQKRYEDQELTGKNLPARFLVDTP 600
 QY 601 GLPNTLFGDVAMVVEFLSCYSGLLLPDAQYPTAVSLMEALSADKGGFLYINRVLLQ 660
 Db 601 GLPNTLFGDVAMVVEFLSCYSGLLLPDAQYPTAVSLMEALSADKGGFLYINRVLLQ 660
 QY 661 TLLQTLQDEIAEDYGEELGKMLSEIPLTLHSVSELVRLCLRRSDVQEESEGSDDDNKDS 720
 Db 658 -LLQTLQDEIAEDYGEELGKMLSEIPLTLHSVSELVRLCLRRSDVQEESEGSDDDNKDS 716
 QY 721 AAFEDNEVQDEFLEKLTSEFPFELTSEFKLQILTLCHRLMTYSVDHMETRQOQSAEL 780
 Db 717 AAFEDNEVQDEFLEKLTSEFPFELTSEFKLQILTLCHRLMTYSVDHMETRQOQSAEL 776
 QY 781 WKERLAVLKEENDKKRAEKQKEMAKNKENGKVENGLGKTDRKKRIVKFEPOVDTEAE 840
 Db 777 WKERLAVLKEENDKKRAEKQKEMAKNKENGKVENGLGKTDRKKRIVKFEPOVDTEAE 836
 QY 841 DMISAVKSRRLAIQAKKERETQEREMKVKLRQAEERIRKHAAAAEAFQSGIAKAKL 900
 Db 837 DMISAVKSRRLAIQAKKERETQEREMKVKLRQAEERIRKHAAAAEAFQSGIAKAKL 896
 QY 901 VMRRTPIGTDNRHNRVWLFSDVPGLFIEKGWVHDSIDYRHNHCKDHTVSGDEDYCP 959
 Db 897 VMRRTPIGTDNRHNRVWLFSDVPGLFIEKGWVHDSIDYRHNHCKDHTVSGDEDYCP 955

RESULT 4
 Q6P1K4 PRELIMINARY; PRT; 808 AA.
 AC Q6P1K4;
 DT 05-JUL-2004 (TrEMBLrel. 27, Created)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
 DE BAZ1B protein (Fragment).
 GN Name=BAZ1B;
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Testis;
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altshul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
 RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
 RA Jones S.J., Marra M.A.;

RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RA Strausberg R.;
RL Submitted (JAN-2004) to the EMBL/GenBank/DBSJ databases.
DR EMBL; BC065029; AAH65029.1; -.
DR InterPro; IPR004022; DDT.
DR SMART; SM00571; DDT; 1.
DR PROSITE; PS50827; DDT; 1.
FT NON TER 808
SQ SEQUENCE 808 AA; 93058 MW; FA06BCE0715CEF74 CRC64;

Query Match 51.9%; Score 4144; DB 2; Length 808;
Best Local Similarity 99.0%; Pred. No. 9.4e-163;
Matches 803; Conservative 1; Mismatches 3; Indels 4; Gaps 1;

QY 1 MAPLGRKPPFVNLDPGEPPFTTPTQEARTRREEYERLERYSERIWTCKSTGSSQL 60
DB 1 MAPLGRKPPFVNLDPGEPPFTTPTQEARTRREEYERLERYSERIWTCKSTGSSQL 60

QY 61 THKEAWEBOEVAELLKEEPFAYWEKLVLEMHVHNTASLEKLVDTAWLEIMTKYAVGEEC 120
DB 61 THKEAWEBOEVAELLKEEPFAYWEKLVLEMHVHNTASLEKLVDTAWLEIMTKYAVGEEC 120

QY 121 DFEVGKEMLVKVIKIHLEKVDDEATEKSDGACDPSDDKNSSQIAQDHQKKEVTV 180
DB 121 DFEVGKEMLVKVIKIHLEKVDDEATEKSDGACDPSDDKNSSQIAQDHQKKEVTV 180

QY 181 KEDGRRSINDRARRSPKLTSLKGERKWAAPPKFLPHKYDVVKLQNEDEKIIISNVPADS 240
DB 181 KEDGRRSINDRARRSPKLTSLKGERKWAAPPKFLPHKYDVVKLQNEDEKIIISNVPADS 240

QY 241 LIRTERPPNKEIVRFIRHNALRAGTGENAPWVDELVKYSLPKSFDFLLDPYKMT 300
DB 241 LIRTERPPNKEIVRFIRHNALRAGTGENAPWVDELVKYSLPKSFDFLLDPYKMT 300

QY 301 LNPSTKRKNTGSPDRKPSKSTDNSSLPNPKLWCHVHLKSLGSPKLVKNSKNSK 360
DB 301 LNPSTKRKNTGSPDRKPSKSTDNSSLPNPKLWCHVHLKSLGSPKLVKNSKNSK 360

QY 361 SPEEHLEENKMWSPNKLHTNPHIPKGPAPKPKGKHSKDKPLKAKGRSGILNGOKSTGN 420
DB 361 SPEEHLEENKMWSPNKLHTNPHIPKGPAPKPKGKHSKDKPLKAKGRSGILNGOKSTGN 420

QY 421 SKSPKGLTPKTKKQMTLLDMAKGTQKTRAPRNSGGTPTTSSKPKHPLPPAALHLIA 480
DB 421 SKSPKGLTPKTKKQMTLLDMAKGTQKTRAPRNSGGTPTTSSKPKHPLPPAALHLIA 480

QY 481 YYKENKREDKRSALSCVISKTARLLSSDRARLPPEELRSLVQRYELLEHKRWASMS 540
DB 481 YYKENKREDKRSALSCVISKTARLLSSDRARLPPEELRSLVQRYELLEHKRWASMS 540

QY 541 EQRKYLKKRBEELKKLKEKAKERREKEMLEKQRYEDELTKGNLPAPFLVDTPE 600
DB 541 EQRKYLKKRBEELKKLKEKAKERREKEMLEKQRYEDELTKGNLPAPFLVDTPE 600

QY 601 GLPNTLFGDVAMVVFSLCYSGLLLPDAQYPIITAVSLMEALSADKGGFYLNRVLVILLQ 660
DB 601 GLPNTLFGDVAMVVFSLCYSGLLLPDAQYPIITAVSLMEALSADKGGFYLNRVLVILLQ 660

QY 661 TLLQTLQDEIAEDYEGELGMKLSLPTLHLSVSELVRLCLRRSDVQEESEGSDTDDNKDS 720
DB 658 -LLQTLQDEIAEDYEGELGMKLSLPTLHLSVSELVRLCLRRSDVQEESEGSDTDDNKDS 716

QY 721 AAPEDNEVDQEFLEKLETSSEFELTSEKQLQILTALCHRLIMTYSVDHMETRQOMSABL 780
DB 717 AAPEDNEVDQEFLEKLETSSEFELTSEKQLQILTALCHRLIMTYSVDHMETRQOMSABL 776

QY 781 WKERLAVLKEENDKKRAEKQKKEAKNKE 811
DB 777 WKERLAVLKEENDKKRAEKQKKEAKNKK 807

DB 777 WKERLAVLKEENDKKRAEKQKKEAKNKK 807

RESULT 5
Q8UVR4
ID Q8UVR4 PRELIMINARY; PRT; 1079 AA.
AC Q8UVR4;
DT 01-MAR-2002 (TRENBLrel. 20, Created)
DT 01-MAR-2002 (TRENBLrel. 20, Last sequence update)
DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)
DE Williams syndrome transcription factor (Fragment).
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
EX MEDLINE=21668941; PubMed=11809820;
RA MacCallum D.E., Losada A., Kobayashi R., Hirano T.;
RT "ISWI remodeling complexes in xenopus egg extracts: identification as
RT major chromosomal components that are regulated by INCENP-aurora B.";
RL Mol. Biol. Cell 13:25-39(2002).
DR EMBL; AF412333; AAL60161.1; -.
DR HSSP; Q9UIG0; 1F62.
DR GO; GO:0000151; C:ubiquitin ligase complex; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0004842; F:ubiquitin-protein ligase activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0016567; P:protein ubiquitination; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR Pfam; PF00439; Bromodomain; 1.
DR Pfam; PF00628; PHD; 1.
DR PRINTS; PR00503; BROMODOMAIN.
DR SMART; SM00297; BROMO; 1.
DR SMART; SM00249; PHD; 1.
DR SMART; SM00184; RING; 1.
DR PROSITE; PS50014; BROMODOMAIN_2; 1.
DR PROSITE; PS50827; DDT; 1.
DR PROSITE; PS01359; ZF_PHD_1; 1.
DR PROSITE; PS50016; ZF_PHD_2; 1.
DR PROSITE; PS50089; ZF_RING_2; 1.
FT NON TER 1
SQ SEQUENCE 1079 AA; 124503 MW; 2AF4689FFFE9E83DF CRC64;

Query Match 46.7%; Score 3728; DB 2; Length 1079;
Best Local Similarity 65.9%; Pred. No. 1.7e-145;
Matches 737; Conservative 140; Mismatches 186; Indels 56; Gaps 15;

QY 333 HIPKGPAPKPKGKHSKDKPLKAKGRSGILNGOKSTGNKSPKGLKTPKTKKQMTLLD 442
DB 1 NLNKGK---SLTGKSKDKKPKNGKSKQ-VLNGQKRVSGKTRSPKCAVKGP--KLQMTLLD 54

QY 443 MAKGTQKTRAPRNSGGTPTTSSKPKHPLPPAALHLIAYKENKREDKRSALSCVISKT 502
DB 55 MAKSTPKVSRQKGGSSSTPRSSSKPKYLPPAALHLISYFRDNKREDKRSALSKV 114

QY 503 ARLLSSDRARLPPEELRSLVQRYELLEHKRWASMSSEQRKEYLKKRBEELKKLKEKA 562
DB 115 ARMLTAEDRKSLPDDLQELVQRYLLEHRLKQWAVMTEQREYMRKKREALKVRKEKA 174

QY 563 KERRREKEMLEKQRYEDELTKGNLPAPFLVDTPEGLPNTLFGDVAMVVFSLCYSG 622
DB 175 RERKQKEERLEKQRYEDELTKGLPTFKLVDTPPEGLPNALFGDVAMVTEFLSGYSD 234

QY 623 LLLPDAQYPIITAVSLMEALSADKGGFYLNRVLVILLQTLQTLQDEIAEDYEGELGMK 682
DB 235 LLLPDAQYPIITAVSLMEALAEKGGFMYLNRVLV---LLQTLQDEIAEDYEGELGMK 290

QY 693 SEIPLTLHLSVSELVRLCLRRSDV---QEESEGSDTDDNKDSAAFEDNEVDQEFLEKLETS 739
DB 291 SEIPLVLHLSSELVRLCLRRKSDSPAGESDSTEKGDSDSVGSVAYQDDEVEDEYLEKLETS 350

QY 740 EPELTSEKLOLITLALCHRIILMTYSVQDHMETRQOMSAELMWERLAVLKEENDKKRAEK 799
DB 351 EPELTSEKLOLITLALCHRIILMTYSVQDHMETRQOMSAELMWERLAVLKEENDKKRAEK 410
QY 800 QQRKEM-EAKNKENGKVGNGTDRKRIIVKPEPOVDTEAEDMISAVKSRRLLATQAKK 858
DB 411 QQRKEGTVPKE-----EVQAAKIVKQB--KINTQODNDAEDMISAVKSRRLLQAWQAKK 464
QY 859 ERIQIREMKVLERQAEERIRKHKAAAEKAFQEGIAKAKLVMRRTPIGTDRHNHRYWL 918
DB 465 EKEHEKLTKERIERETEERESGRKQKASAEKAFHEGIAKAKLVRLSPGLTDRHNHRYWL 524
QY 919 FSDVEPGLFIEKGVWHDSDIDRFNHHCKOHTVSGDEDY-----CPSKXANLQGNAS 970
DB 525 FSDVEPGLFIEKGVWHDSDINRFSPSKQDSEQDAESEDANSSIGCPDDSTQREKHA- 583
QY 971 MNTQGTATEVAVETTPKQGNLWFLCDSQKDELNCLHPQGIRESOLKLEREKYQ 1030
DB 584 -----ETTPKQGNLWFLCDSQKDELNCLHPQGIRESOLKLEREKYQ 630
QY 1031 DIHSHILARKPNLGLKSCDGNQELNLFRLSDLIEVATRLQKGLGYVBETSEFEARVIS 1090
DB 631 DIMHSHILARKPNLGLKSCDGNQELNLFRLSDLIEVATRLQKGLGYLDTTFEFAKVT 690
QY 1091 LEKLDGFGCVIALQASVTKFLQGFMAPKQKRKLQSD-SAKTEEVDEEKKQVVEAKV 1149
DB 691 FENLKDGFGEICVFLQAIVIKFLQGFMAPKQKRKHQSBEEAAKAEQDEEKKMABEAKV 750
QY 1150 ASALEKWKTAIREAOTFSRMVLLGLMDACIKWDSNAENARCKVCPKKGDDKLLCDSK 1209
DB 751 ASAVEKWKVAIRDAOTFSRMVLLGLMDACIKWDSSENAENARCKVCKRKGDDKLLCDSK 810
QY 1210 NXAFHFLCLRPALYVDPGEWCPACQATARNRSGRNYTESASEDSEDEDEE 1269
DB 811 NXAFHFLCLRPVLFNIPDGEWCPACQATARNRSGRNYAESDQDEDEESEE 870
QY 1270 EEEE--EEEDVEVAGLRPRKRTIRKHSVIPPAARSGRRPKKPHSTRSQKAPPVD 1327
DB 871 EEESEDEEEQBMQRLSRKAAGKPG--RPTRRGRPKNNTHS-RVSRQRVDE 926
QY 1328 DAEVDELVLQTRSSRSROSLQKCEILHKIVKRFSPREPVTTRDAEYDYVITHP 1387
DB 927 EADVEEMVQSKPTSRQNOEQFKCEILAKLKYRFPFPFNADEIETKVVVTP 986
QY 1388 MDPQTQVQKSCGSRVSOEFTDMQVFTNAEVNCRGSHVLSQVTKTEQCLVLLKH 1447
DB 987 MDPQTQVQKSCGSRVSOEFTDMQVFTNAEVNCRGSHVLSQVTKTEQCLVLLKH 1046
QY 1448 LPHGPVVRKRKKFPDRLAEDBGDSPEAVGSRDEDR 1486
DB 1047 LPAHTYQRRHK-----HQSPPEPETANPGRGRKQK 1078
RESULT 6
Q8CAU9 ID Q8CAU9 PRELIMINARY; PRT; 657 AA.
AC Q8CAU9
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Mus musculus 16 days neonate thymus cDNA, RIKEN full-length enriched
DE library, clone:A130044E01 product:bromodomain adjacent to zinc finger
DE domain, 1B, full insert sequence. (Fragment).
GN Name=Baz1b;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Thymus;
RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA Carninci P., Hayashizaki Y.,

RT "High-efficiency full-length cDNA cloning."; Meth. Enzymol. 303:19-44(1999). [2]
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Thymus;
RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RA RIKEN FANTOM Consortium;
RT "Functional annotation of a full-length mouse cDNA collection."; Nature 409:685-690(2001). [3]
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Thymus;
RA The FANTOM Consortium;
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs."; Nature 420:563-573(2002). [4]
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Thymus;
RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M., Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y., Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K., Fujisake S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M., Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J., Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer."; Genome Res. 10:1757-1771(2000). [6]
RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Thymus;
RX MEDLINE=2030913; PubMed=11076861; DOI=10.1101/gr.152600;
RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P., Konno H., Akiyama J., Nishi K., Kitsumai T., Tashiro H., Itoh M., Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Hazada A., Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K., Fujisake S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M., Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J., Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer."; Genome Res. 10:1757-1771(2000). [6]
RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Thymus;
RX MEDLINE=2030913; PubMed=11076861; DOI=10.1101/gr.152600;
RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P., Konno H., Akiyama J., Nishi K., Kitsumai T., Tashiro H., Itoh M., Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Hazada A., Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K., Fujisake S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M., Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J., Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer."; Genome Res. 10:1757-1771(2000). [6]
RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Thymus;
RX MEDLINE=2030913; PubMed=11076861; DOI=10.1101/gr.152600;
RA Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P., Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W., Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T., Hori F., Inotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T., Katoh H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S., Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M., Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y., Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H., Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M., Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T., Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK037737; BAC29862.1; --
DR HSSP; OSUIG0; 1F62.
DR MGD; MGI:1353499; Baz1b.
DR CO; GO:0005721; C:centric heterochromatin; IDA.
DR GO; GO:0000793; C:condensed chromosome; IDA.
DR GO; GO:0005634; C:nucleus; IDA.
DR GO; GO:0005515; F:protein binding; IPI.
DR GO; GO:0006333; P:chromatin assembly or disassembly; IDA.
DR GO; GO:0006338; P:chromatin remodeling; IDA.
DR InterPro; IPR001487; Bromodomain.
DR InterPro; IPR011011; FYVE_PHD_Znf.
DR InterPro; IPR001965; Znf_PHD.
DR InterPro; IPR001841; Znf_PHD.
DR Pfam; PF00439; Bromodomain; 1.
DR Pfam; PF00628; PHD; 1.
DR PRINTS; PR00503; BROMODOMAIN.
DR SMART; SM00297; BROMO; 1.
DR SMART; SM00249; PHD; 1.

Q9V9T4;
01-MAY-2000 (TrEMBLrel. 13, Created)
01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
CG1966-PA.
Name:Acg1; ORFNames=CG1966;
Drosophila melanogaster (Fruit fly).
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
NCBI_TaxID=7227;
[1]
SEQUENCE FROM N.A.
MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;
Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
Sutton G.G., Wortman J.R., Yeandle M.D., Zhang Q., Chen L.X.,
Brandon R.C., Rogers Y.H., Blazer R.G., Champe M., Pfeiffer B.D.,
Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
Aller J.F., Bagayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
Beeson K.V., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
Borkova D., Botchan M.R., Bouck J., Brockstein P., Brothier P.,
Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
de Paulos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
Nelson D.R., Nelson K.A., Nixon K., Nusser D.R., Pacleb J.M.,
Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
Wang Z.Y., Massarman D.A., Weinstein G.M., Weissenbach J.,
Williams S.M., Woodgett, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu H.O.,
Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
"The genome sequence of *Drosophila melanogaster*.";
Science 287:2185-2195(2000).
[2]
SEQUENCE FROM N.A.
MEDLINE=42426065; PubMed=12537568;
Celniker S.E., Wheeler D.A., Kronmiller B., Carlson J.W., Halpern A.,
Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,
George R.A., Hoskins R.A., Laverty T., Muzny D.M., Nelson C.R.,
Pacleb J.M., Park S., Pfeiffer B.D., Richards D., Sodergren E.J.,
Svirskas R., Taber P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
Weinstock G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.;
"Finishing a whole-genome shotgun: Release 3 of the *Drosophila*
melanogaster euchromatic genome sequence.";
Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).
[3]
SEQUENCE FROM N.A.
MEDLINE=42426070; PubMed=12537573;
Kaminker J.S., Bergman C.M., Kronmiller B., Carlson J., Svirskas R.,
Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,
Ashburner M., Celniker S.E.;
"The transposable elements of the *Drosophila melanogaster* euchromatic
genomes perspective.";
Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).
[4]

Query Match	8.3%	Score 666;	DB 2;	Length 1476;
Best Local Similarity	18.4%	Pred. No. 2.5e-19;		
Matches 331;	Conservative 256;	Mismatches 504;	Indels 710;	Gaps 56
Qy	3	PLLRKPPPLVNPPLGEEFF-----FTIPHTQEAFTREERYEARLERYSERIWTCKSG	56	
Db	2	PICKREGDLNKGKGNKTFHDNDQVFCYITKRIFRDYEHYFRHVMVINSTVMQCEATG	61	
Qy	57	SSQLTHKEAWBEEQBEAELLKEEPFAMVEKLVLEMVHHTASLEKLVDTAWLEIMTK-YA	115	
Db	62	KENLTYEAVKSER--AARKWEQFKQLSRSPVLLVVEHAQQSAVNTLWIVAKFLKRVF	120	
Qy	116	VGECDPFGVGEKMLKVKIVKHIPLEKVDDEATEKKSDGACDPSDSENSQIAQDHQK	175	
Db	121	IGEEVSVQAKKNATYVLGVKL-----DKNMPPLNGIYED	156	
Qy	176	KETVVKEDBGRRESINDRARSPRKLPTSLKGERKWAPPKPLPHKYDVKLQNEDKIISN	235	
Db	157	TDNLVY-----RURPNKGDPSAE-----LD	176	
Qy	236	VPADSLIRTERPPNKEIVRYFIRHNALRAGTGENAPVWVEDLVKYSLPKSFSDFLDLP	295	
Db	177	LPFQLRRSRMEFNLNLSMFIKSNVSRV-----DGLLR--PKPEAYKQVTDPD	223	
Qy	296	YKYMTLNPSTRKNTGSPDRKPSKSKTDSNSSLPLNPKLVHVKLSLGSGLKVNK	355	
Db	224	---GVNFST-----IFIGMPRYSPAKIK--	244	
Qy	356	SKNSKSPBEHLEEMKMKMSPNKLHTNFHFKGPPAKPGKHSIDPLKAKGRSKGILNQ	415	
Db	245	-----KPDGK-----KQSTLNKYIVAGE	262	
Qy	416	KSTGNSKSPKGLKTPTKMQQMTLLDMAKGTQKMTAPRNSGGTPRTSSKPKHKLPPAA	475	
Db	263	ATAAKSKAKA-----	273	
Qy	476	LHLIAYYKENKDREDKRSALSCVISKTARLLSSEDRARLPBELRSLVQKRYELLEHKRW	535	
Db	274	-----SDAKSLAELELV-----	286	
Qy	536	ASMSEBORKEYLKKRBEELKKLKEKAKEREKEMLEKQ-----KRYEDQELTG-KN	589	
Db	287	-----KREKAELIELEKQAEKKAQILIERVENECNLLQKTDLDLERTDQKV	333	
Qy	590	LPAFLVDTPGLNPTLFGDVAMVVFELSCYSGLL--LPDAQYPITAVSLMEALSA-DKG	646	
Db	334	LPYRQIVTL--LPEHLUGDAFMRFREFMHTYTGLLSGIEVFQNLSFYEMTRALTARETA	391	
Qy	647	GFLYNRLVLILOTL--LOTLLODEIAEDY-----GELG	679	
Db	392	G--PLSDILLVLLGTVPDLOKEBEECAVTYLDRAAQTOEPYWSAAQAKSHLYAKRHFS	449	
Qy	680	MKLSIEIPLTHSVSELVRLCLRRSD--VOBESG-----SOTDDNKDSAAAFEDNEVQDEF	732	
Db	450	FKVNELPLDALTLNEVLRHLHLSGAFVNEKAERWRVMYRNGYSSKEDPGLELRLEHSHI	509	
Qy	733	LEKLETSEFPELTSSEKLOIILTALCHRILMTYS-VODEMETRQQSABELWKERLAVLKEE	791	
Db	510	LRILKNHSVQLKPKDINMLIRCLMSQI-MTYSGTINLIEERMEQATAKARQDLRALVUGE	568	
Qy	792	NDKKRAEKQKRWKAKNKENGKVLGKTDKRIKVPFPQVDTEAEDM--ISAVKSR	850	
Db	569	NKRLAAVEINRKL-----TQHHLEVNGVEPEK	597	
Qy	851	LLAIOAKREBIEQREMVKVLERQAAEBEIRIKHKAABEAKFQEGIAKALVMRRTPIGTD	910	
Db	598	REALVEKLKKSIAE-----LHAQSDQOH-RKHELQMLK-----LHSQLFNFLVYLGWD	644	
Qy	911	RHNRYWLFSDVEPGLFIEKGWHDSDIYRFNHCKDHTVS-----	951	
Db	645	RCRYKYTVL--ESMPGIIFVHS--PDSDL-----TCLEQPIFNKSQIETRQOSALPKNRKD	696	

Qy	952	-----GDEDYCPRSKKAANLGNASMTQH-----GTATEVAVETTP-----	988
Db	697	LRVYLLKLYGDDEK-KTKKAKHSLNENKQEHRLNGASBPMVDSPAPTHFELLMC	755
Qy	989	-----KOGO-NLWFLCDSQKELDELINCLHPQGIRESQL-----	1021
Db	756	SGDKRSCIVHDSRMGQQRWAYIYKAEIDELIKALNPGLREYELLQELSVLSLIEQH	815
Qy	1022	-----KERLEKRYODIHSIHLARK-----PNLGLKSCDGNQOELLNF-----	1058
Db	816	AKTCPVDLLSLENETMKFMAAMES-ETNRKYGEANFGLPNGTDLNEVMRLHVDRIIQ	874
Qy	1059	-----LRSDLI-----EVATRLQKGLGVVE-----ETSE-1083	
Db	875	FENDIYTGDLGRKLVKDMEKWRSDLLGNGYDAQCKLQWGPCKLEDAGSNDNESHETHEE	934
Qy	1084	-----FEARVISLEKLKD-----FGECVIAL1104	
Db	935	DGALLGKYARKPYRDPGMYTJASADTKPLPDSDDBEDQHTNAVSPIAVHNWASALLQV	994
Qy	1105	QASVIKKFLQ---GFMA--PKQKRKLQSDSAKTEEVDEBKVMVEAKVASALEKWKTA	1159
Db	995	EQAIKGFLEKPEYGMKKWDPKQEAALKACD-----SRHQWEVS1033	
Qy	1160	IREAQTSRMHVLLGLMDACIKWDMASENARCKVCPKKGEDDKLLCDECNKAFHLPCLR	1219
Db	1034	LMESTSPAQVFLHNLILHDCIQWRRSTNKSCLKVCRRGSDPEKMLLDCDECNAGTHMPCLK	1093
Qy	1220	PALYEPDGEWQCPACOPATARNRSGRNYTESASE-----DSEDDSEDEEEEEEE	1273
Db	1094	PKURSVPPGNWYCNDCVKSGLGNGQNEKKQATKKRKFIVEEDDEATDEEBEKKD	1153
Qy	1274	EEBEDYVAGLRPRKTIIRKHSV---IPPAARSG---RRPGKKPHSTRRSQPK-----	1322
Db	1154	DDMTDEDAHENEKHEDVZDSVSTPSSSRVNGRILARRPTRTSRLTSKTEEHA	1213
Qy	1323	-----APVDDAEVDELVLQT-----	1338
Db	1214	QEDVDSGDVSDDASLTAGETIEDSEDEKVCQCFYDGGEIKVCQCLFFHLECVHLKR	1273
Qy	1339	-----	1338
Db	1274	PPRTDFVCKTKPMQPPRRRHSNMNGDHRDEEPAKPRNRSRLCIDKTARPSNGNN	1333
Qy	1339	-----KXSSRRQS---LELQKCEIILHKIVKRYFSPFPREPVTTRDEAEDY	1380
Db	1334	NNNNNSSVNNNHRSGRTNHEMPLNSAALYDLLEQIMKHAAPFLRPVLTSEVPDY	1393
Qy	1381	YDVITHPMDQTVQNKSCGSYSRVQFFLMDKQVFTNAEYVNCGRSHVLSCMVKTEOCL	1440
Db	1394	HQIIKTPMDLAKIKSKLWGMAYQLNEELLSDIQLVFRNCDLYNVEGNEIYDAGCQLERFV	1453
Qy	1441	V1441	
Db	1454	I1454	
RESULT 11			
QYOWI	ID	QYOWI1	PRELIMINARY; PRT; 1476 AA.
AC	AC	QYOWI1	
DT	01-NOV-1999	(TReMBLrel. 12, Created)	
DT	01-NOV-1999	(TReMBLrel. 12, Last sequence update)	
DT	01-MAR-2004	(TReMBLrel. 26, Last annotation update)	
DE	ATP-dependent chromatin assembly factor large subunit.		
GN	Name=Acfl;		
OS	Drosophila melanogaster (Fruit fly).		
OC	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;		
OC	Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;		
OC	Ephydroidea; Drosophilidae; Drosophila.		
OK	NCBI_TaxID=7227;		
RN	[1]		
RP	SEQUENCE FROM N.A.		


```
QY 1339 -----KRSRRS-----LELQKCEILHIVKVFREFPVR 1374
Db 1328 PSNGNNNNNNNSVNNHRRGRRTNEHPLNSAALYDLLEQIMCHKAAPFLREVL 1387
QY 1375 DEAEYDVITHPMDFTQVONKSCGYSRVQEFITDMKQVETNAEVYVNCRGSHVLS 1434
Db 1388 SEVPDHOIITPMDLAKIKSLNMGAYQLNEELSDIQLVFRNCDLYNVEGNEIYDAGC 1447
QY 1435 KTEQCLV 1441
Db 1448 QLERFVI 1454

RESULT 12
Q7PRP9
ID Q7PRP9 PRELIMINARY; PRT; 1357 AA.
AC Q7PRP9;
DT 01-MAR-2004 (T-EMBLrel. 26, Created)
DT 01-MAR-2004 (T-EMBLrel. 26, Last sequence update)
DE ENSANGP0000001532 (Fragment).
GN Name=ENSANGG00000001297;
OS Anopheles gambiae str. BEST.
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Anopheles.
OX NCBI_TaxID=180454;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PEST;
RA Anopheles Genome Sequencing Consortium;
CC Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AAA01008847; EAA06826.2; -.
DR HSSP; Q9UIG0; 1F62;
DR GO; GO:0003677; F1DNA binding; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR001487; Bromodomain.
DR InterPro; IPR004022; DDT.
DR InterPro; IPR011011; FYVE PHD ZnF.
DR InterPro; IPR001965; ZnF_EHD.
DR Pfam; PF00439; Bromodomain; 1.
DR PRINTS; PR00503; BROMODOMAIN.
DR PROSITE; PS00633; BROMODOMAIN 1; 1.
DR PROSITE; PS50014; BROMODOMAIN_2; 1.
DR PROSITE; PS50827; DDT.1.
DR PROSITE; PS50016; ZF_PHD_2; 1.
DR NON_TER 1357 1357
SQ SEQUENCE 1357 AA; 153055 MW; D42FD563C34F5287 CRC64;

Query Match 8.2%; Score 655; DB 2; Length 1357;
Best Local Similarity 19.8%; Pred. No. 6.5e-19;
Matches 337; Conservative 226; Mismatches 447; Indels 690; Gaps 64;

QY 18 GEEFFFTIHTQAFRTREYEALRERYSERIWTCKSTGSSQLTHKEAWEEEOEVALLK 77
Db 21 GDEVLCQP-TGIFSNYEDYFHRIMLSSWWTCAWTPNLTLYEALDSEKDKARKLK 79
QY 78 EEPFANYEKLVLDMVHHNTASLEKLVDTAWLEIMTKYAVGEECDPEVGKMKLVKIVKI 137
Db 80 -TPPAAVKGFLLVASH----- 95
QY 138 HPLEKVEATEKKSQDAGCDSPSSDKENSQIAQDHQKQKVVKEDEGRRESIN--DRAR 195
Db 96 -----TQRTS-----INEMHEDYVGFYKDLHFKGESVDAMDPR 129
QY 196 RSPRKLTSLKKGKRWAPPKFLPHKYDVKLQNEDKIISNPADSLIRTERPPNKEIVRY 255
Db 130 KAFRR-----ATTINLVDSLT-----TPN--ALTY 153
QY 256 FIRHNALRAGTGENAPWVVEDELVKYSLPSKPSDFLLDPYKYMTLNPSTKRKNTGSPDR 315
```

```
Db 154 QVR-----ABDDMT-----PKVMTVLPGNVKR-----DR 177
QY 316 KPSKSKTDNSSLSSPLNPKLWCHVHLKSLSGSP---LKVKNKSNKSPSEHLEEMKMM 372
Db 178 SALTREK-----CKLFLKHVEAGPGQLRIKADSLARF----- 211
QY 373 MSPNKLHTNPHIPKKGPPAKKPGKSDKPLKAGRSKG-ILNGQKSTGSKSPKKGKLT 431
Db 212 -----VTGEGWTDGLVFHGTGTP---SFQQSKRLK-- 237
QY 432 KTYMKQWTLTLDMAKGTQKMTAPRNSGGTPRTSSKPHKHLPPAALHLIAIYKKNOREDK 491
Db 238 -----QNAERGEAGGG----- 249
QY 492 RSALSCVISKTARLSSSEDRARLPEELRSLVQKRYELLEHKKRWASMSBQRKEYLAKKR 551
Db 250 -----AAASTALEAKAKKQEBEAAQLRURVE-----QALAKRR 283
QY 552 EELKKKLKKEKAKERREKEMLERLEKQ-----KRY-----EDQELTGKN-LPAFLRLVDTPEG 601
Db 284 ABEKAL-----LAQVVLAKKYNAVLEDOELPDQRPPLPVPVRPL-- 326
QY 602 LPNTLFGDVAMVVFELSCYSGLLLPDAQVP--ITAVSLMEALSADGGFLYLNRLVILL 659
Db 327 IAAHFSSVFVILEYLNFPADLIARSKFPGLTIHLLEALILRE-----VNGPLSDIF 381
QY 660 QTLTQTLQDETAEDYGE-----LGMKLESEIP 686
Db 382 QVLLSAIFTQEEENEEESVRYERIEGLAQKQTVPEQVRAADAALWCEKHYSKSTSEMP 441
QY 687 LTLHSVSELVRL--CLRRSDVQEESEGSDTDNDKDSAAFEDNE---VQD--EFEKLETS 739
Db 442 IDSMTVSELLRLHLFLASGALVEEARAHRQYQNGGYGSGDDPGLRLVRDYPHILRALNWY 501
QY 740 EPELTSEKQLQILTALCHRLIMTYSVQDHMETRQOMS-----AELWKEHLAVLKEE 791
Db 502 SVYQLPVGDIITQLCCLIHQLLTYSYVRDLVEERVEKARTARTSYQANRWAQRLTVKA- 560
QY 792 NDKKRAEKQKREMEAKNKENGKVGKLTDRKKRIVKFEPOVDTEAEDMISAVKSRRL 851
Db 561 -----GSLKNA-ARDMMKRELAFEGELAREE-----YRRK 591
QY 852 LAIQ-----AKKERIEQEREMKVKLERQAEERIRKHKAAAEKAFQEGIAKAKLVWRTP 906
Db 592 LTRDLDEQCARLEADAQ-RQLKV---LQAESERLK-----EDFFDYQIY----- 631
QY 907 IGTDRNHRNYLWLFSDDEVPGLFTEKGWVHDSIDYRFNHHCKDHT---VSGDEDYCPRSKKA 963
Db 632 LGTDRCYRNYLWF-ESLPLGLFVEH-----DRTYAGRCILDRPTPHIPGLAACAPQORKK 683
QY 964 NL-----GKNVSMNTQH-GTATEVAVETTTPKQGNLWFLCDSQ----- 1001
Db 684 YITRAIMRCAGTGVVALLDADAIAADANAVSSSSTGPAFVEEELMTAQDPDPCPVHTEQ 743
QY 1002 -----KELDELINCLHPGIGRESOLKERLEKRYQDIHSHIHLARKP--NLGL 1046
Db 744 HPCTVCGWGYATADELDALIRSNARGAREKQLRETLECE-RDLIYV-HIAKPLDKLSV 801
QY 1047 KSCDGNQELLNFLRSLDIEVATRLQKG-----GLGYVEETSEFEARV 1088
Db 802 GEAD-----RAGALADIASRNQRYDAPNPNFSGPTGPEPNEILEAVFLENLELEAKI 853
QY 1089 -----ISLEKLDGFGECVIALQASVIKFLQGFMA-----PKQRRKQLQSEDAKTE 1135
Db 854 TVGYLGWVMVRDRDKWREALEA-----HGYDAQOTDRFLVMGPKRLAPKAEEGDNGREQ 906
QY 1136 EYDEEKKKMWEE----- 1146
Db 907 QDGEQKQVQKQENGPNDRVDSQDQKPVVAKLPGCGDLDSERDSLRLMSEAVQSLGRA 966
QY 1147 -AKVASALE-----KWKTAIRAQOTFSRMHVLL 1173
```

Db 967 LQVACIDPKFLRHPFGPKDKDKAVNQKQHEGLKNLVRWESVLMQSTCYAQLFLHY 1026

Qy 1174 GMLDACIKWDMAEARCKVKPKGEDDKLILCDECNKAFHLFCRLPALYVPGSEWQCP 1233

Db 1027 NVLYDAIHSRGAERICCMICRRKGGDPGLTLLCDECNKACHTYCLPKLKEYPAGDWYCK 1086

Qy 1234 ACQP-----ATARRNSGRNNT-BESASEDSEDDSEEEEEEEEEDEYEVAGLRL 1286

Db 1087 RCRPENFEKPSANRKKKKMFSGSDGAEENDDGSDADYRAKSKMLVKVAKVA-KRS 1145

Qy 1287 RPR-----KTR-----GRHS 1297

Db 1146 KPKAKAKPTIVTSTGRSTSTAATNGAVRLALERWRRSEAOQCPAAVHQEAQAGGGD 1205

Qy 1298 VTPPAARSRRPKKPHSTRSQPKAP-----PV-----DAAVDVLVQTKRSSR 1343

Db 1206 VSPFASKARRA-----TNGRAKVGPGGTIDSEEPFLGGGGTGDPLNSVALYT----- 1255

Qy 1344 RQSLQKQCEETLHKIVKVFSGWPPREPVTTRDEAEDYDVITHPMDFOTVQNKCSGSYR 1403

Db 1256 -----JIDILKHPNSWPNRPVSAKEVPDYAVIKSPMDFAIKSKLNNGDYK 1304

Qy 1404 SVQEFITDMKQVFTNAEVN 1423

Db 1305 INEQMLSDVQLVFRNCDLYN 1324

RESULT 13

Q9NSL9 PRELIMINARY; PRT; 1427 AA.

AC Q9NSL9;

DT 01-OCT-2000 (TReMBLrel. 15, Created)

DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)

DT 01-MAR-2004 (TReMBLrel. 26, Last annotation update)

DE Hypothetical protein H20J04.2.

GN Name=H20J04.2; ORFNames=H20J04.2;

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;

OC Rhabditidae; Feloderinae; Caenorhabditis.

OX NCBI_TaxID=6239;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Bristol N2;

RX MEDLINE=99069613; PubMed=9851916;

RG WormBase Consortium;

RT "Genome sequence of the nematode C. elegans: a platform for

RT investigating biology. The C. elegans Sequencing Consortium.";

RL Science 282:2012-2018(1998).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=Bristol N2;

RA Dauphin S., Antoniou B., Gibson A.;

RT "The sequence of C. elegans fosmid H20J04.";

RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN=Bristol N2;

RA Waterston R.H.;

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

RN [4]

RP SEQUENCE FROM N.A.

RC STRAIN=Bristol N2;

RA Waterston R.;

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

RN [5]

RP SEQUENCE FROM N.A.

RC STRAIN=Bristol N2;

RA Waterston R.;

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

RN [6]

RP SEQUENCE FROM N.A.

RC STRAIN=Bristol N2;

RA Waterston R.;

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.

RN [7]

RP SEQUENCE FROM N.A.

RC STRAIN=Bristol N2;

RA Waterston R.;

RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.

RN [8]

RP SEQUENCE FROM N.A.

RC STRAIN=Bristol N2;

RA Waterston R.;

RL Submitted (OCT-2002) to the EMBL/GenBank/DBJ databases.

RN [9]

RP SEQUENCE FROM N.A.

RC STRAIN=Bristol N2;

RA Waterston R.;

RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.

RN [10]

RP SEQUENCE FROM N.A.

RC STRAIN=Bristol N2;

RG WormBase Consortium;

RL Submitted (SEP-2004) to the EMBL/GenBank/DBJ databases.

DR EMBL; AC006661; AAF39888.2; -.

DR HSSP; Q03330; IE61.

DR WormBase; WBGene00019217; H20J04.2.

DR WormPep; H20J04.2; CE27187.

DR GO; GO:0005634; C:nucleus; IEA.

DR GO; GO:0003677; P:regulation of transcription, DNA-dependent; IEA.

DR GO; GO:0006355; P:dna replication; IEA.

DR InterPro; IPR000637; A+T hook.

DR InterPro; IPR001487; Bromodomain.

DR InterPro; IPR004022; DDT.

DR InterPro; IPR001965; Znf_PHD.

DR Pfam; PF02178; AT_hook; 1.

DR Pfam; PF00439; Bromodomain; 1.

DR Pfam; PF00628; PHD; 1.

DR PRINTS; PR00503; BROMODOMAIN.

DR SMART; SM00384; AT_hook; 1.

DR SMART; SM00297; BRÖMO; 1.

DR SMART; SM00571; DDT; 1.

DR SMART; SM00249; PHD; 1.

DR PROSITE; PS00633; BROMODOMAIN_1; 1.

DR PROSITE; PS00014; BROMODOMAIN_2; 1.

DR PROSITE; PS00827; DDT; 1.

DR PROSITE; PS01359; ZF_PHD_1; UNKNOWN_1.

DR PROSITE; PS00016; ZF_PHD_2; 1.

QY Hypothetical protein.

SQ SEQUENCE 1427 AA; 165227 MW; 401039F8C2BA4D3 CRC64;

Query Match 8.2%; Score 654.5; DB 2; Length 1427;

Best Local Similarity 20.3%; Pred. No. 7.3e-19;

Matches 334; Conservative 248; Mismatches 574; Indels 493; Gaps 61;

QY 3 PLLGRKPF---PLVNPLGEEPPFTIPHTQEAFTRTREYEARELRYSERIWTCKSTGSSQ 59

Db 17 PLLHKQPEPQSPVDPGVTDRTPVFYCKATKEVFLTHEYFENRMLINSTAWSCSLTRKSN 76

QY 60 LTHKEAWEEEQVAVELLKEEPPAWYKLVLEMVHH--NTASLEKLVDTAWLEMTKYAVG 117

Db 77 LTYFEAIAASERAEQELC-NFPTALEIPTILIVHKYTNRGRFEDLVNDIYHLIKORFFNN 135

QY 118 EECDFEVGKMKLVKVIKIHPLKVDDEATEKKSDGACDSSPSKENSQIAQDHQKKE 177

Db 136 EEVAYAEKSRKFL-AKIVESNQIEE-----QNFSF--EGGSEL-----KE 172

QY 178 TVVKEDEGRRESINDRARRSPKLPSTSLKKGKRWAPPKFLPHKYVDVKLQNDKLIIS---- 234

Db 173 PIL-----PP-----PDTFRYTLQILDPAVSEEE 196

QY 235 ----NVPADSLIRTERPPNKEIVRYFIRHNALRAGTGENAPWVEDELVKYKSLPSKESD 290

Db 197 MYREGIAFDRLPRSKNIGSRKIRLFLKNCCOMPDS-----RVLTKQFLUE 244

QY 291 FLIDPYKYMTLNPSTKRKNTGSPDRKPKSKKSTDNSSLSPLNPKLWCHVHLKKSLSGSP 350

Db 245 KIDNLY-----WTDVMSGAEPICPQ----- 264
QY 351 LKVKNSKNSKSPBEHLEEMKMMSPNKLHTNFHPKKGPPAKPGKHSDKPLKAKGRSKG 410
Db 265 -----TPALQGRAPNTLKAGDHERKE-----KPKKEERDP-NAPPRPG 304
QY 411 ILNGKSTGNSKSPKGLKTPKTKMOMTLDDMAKGTQKMTAPR-----NSGGTPRTS 464
Db 305 -----RPF-----KTPQKALDKKQKMKRAEAGREAPKDDPDLTSSGA-MTL 348
QY 465 SKPHKLPPAALHLIAIYKENKDREDKRSALSCVISKTARLLSE-----DRARLPE 516
Db 349 SIP-----VSPA-----KKKIRRESGSMG--KKTKILRQOQELCFFFEARRLGI 394
QY 517 ELRSLVQKRYELHKKRWMASMBEORKEYLKKRBEELKKKKEKRAERKEKMLRLEK 576
Db 395 DVGLEQEE-KLLSGK-----AVAD-----FKQVKEKDAEKERAKEKKRKREKTAY 443
QY 577 QKRYED---QELTGKNLPAPRLVDTPEGLPNTLFGDVAMVVEFLSCYGLLLPDAQYPIT 633
Db 444 NKKREDLICNDL--KELPRPKLEIPEWISNAEFEDYLFIFQFNGFKQL----- 492
QY 634 AVSLMEALSADKGGFLY-----NRVILVILQTL--QTLQDEIAEDYGE----- 678
Db 493 --PLKEIRGSEVQFSDIIIAIKNDPQNSFADLLRVLSTRTDIADDEGDEADINNR 550
QY 679 -----GMLKSEIPLTLHVSSELVRLCLRR 702
Db 551 EEVYLINAQNDPAVTHGDSIRDSLHLFKIRKHGKSVRHLFVDMWLTVELVRLIFET 610
QY 703 SDVQSESESDTDNK-----DSAAFEDNEVQDEFLEKLETSEFFELTSEKLOI 752
Db 611 SGYI---TGWATHRHLRYARGNFRGYEDPAYEPTRHPGIMEKRLTLTVFDEAPERLEI 667
QY 753 L7ALCHRLMTYSVQDHMETROQMSAELMKERLAV-----LKEENDKKRA----- 797
Db 668 VKTLIYQLLTYSKFRGHLEQORQNELVELKREQKKAWDVQGEANEARLLLELAPAGA 727
QY 798 -----EKQKREMEAKNKGENV-----GLGKTDKRRKIVKPEQVDTEAEDMISAVKS 848
Db 728 SGAPDVKEQEPVPRRLKAHKAINEGRYKEDLDAIVLESVPVASISLBEIVTA--- 784
QY 849 RLLAIQAKKEREIQREMKVKLEROAEEERIRKHAABKAFQ---EGIAKAKLVNRRTP 906
Db 785 -----RELQKSEFKILMD-----SLISKMFQIYSKISDIRL----- 815
QY 907 IGTDRHNRNYLPSDEVPGLFTEKGM-----VHDSIDYRPNHCKDHTVSGD 953
Db 816 -GSDRAYRY-IVMENLSAILVETATFQELGIYCDPSIIDDPNLNLENEHQEVFICTGN 873
QY 954 EDYCPKRSKANLGKNAASMTQGTATEVAVETTPKQGNLWFLCDSQKELDELLNCLHP 1013
Db 874 MTC-----RVHGDSE-----NNWTRWSYIRDRQEQFOLLKSLNP 908
QY 1014 QGIRESQLKERL-----EKRYQDIISHLARKPNLG-LKSCDQNG 1053
Db 909 RGNREVELLEELNEYPSSLLEILETERLHEEEDENKSWQPMTDNPNPGDTYNIDWA 968
QY 1054 ELLNFRSLDIEVATLQGGLYVE---ETSEFEAR-----VISLKLKDFGEC 1100
Db 969 E-----MRDLLDFEEDIDQMGMSIEKIFECNRIEWRDLNKBGNGVCMLLNEDISIFGEI 1024
QY 1101 VIALQAS-----VIKFLQGFMAPKQRRKQJSDSAKTEEVEDEKKMVEEAKVASALEK 1155
Db 1025 SVNLESEFSPMSUKLAIFAYMIKSIHLKF-----IKAPYISPNDKHNGLKPSSELFTR 1079
QY 1156 WKTATREAOFTSRMHVLLGLMDACIKWDMASNAENARCKVCPKXGDDKLILCDECNKAFHL 1215
Db 1080 WQRALLECESHSAISLFIITFESSIKWDKSRLOQKCRSRRKAAAHDLVLCSEDCNCHYL 1139
QY 1216 FCLRPALYVPGEWQCPACQAPATARNRGRNYTEESASEDSDESEDEEBEE--- 1272

Db 1140 KCAKLDVNSDAPADWMCVTCR-AQQRK-----VENEAKRMARDDDLAESSQEDVAI 1191
QY 1273 ---EEEBEYEVAGLRRLRPRKTI-RGKHSVIPPAAAR--SGRRPK-----KPH 1314
Db 1192 GNISISDNSETAPKTAEPAESCSPTTTSKEPTIRTASGRAVRRVHYSIHGELSUNKR 1251
QY 1315 STRRSQPKAP---PVDAAEV-----DELVLQTKRSRRSQSLBQK----- 1351
Db 1252 KNSGMPVTPSPERPVRNVSVRIPDVENENVLTDDEDDSGENTRKRKTPTTSKSVTSTPT 1311
QY 1352 -----CEIILHKIVKRFSPRPVTRDEAEDYDVITHMDFOTVQ 1394
Db 1312 NDISRVIRNIKEKMTLIETLLKEAMRQBCSWFQFVDSKEYPDYVDYVIKRPMNLTMM 1371
QY 1395 NKSCCSGYSRVOEFLTDMKQVFTNAEAVYN 1423
Db 1372 NKIKQRIYNKPIEVRNDFQLILNSCETYN 1400
RESULT 14
Q6PID9 PRELIMINARY; PRT; 892 AA.
AC O6PID9, PRELIMINARY; PRT; 892 AA.
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DE BC065123 protein (Fragment).
GN Name=BC065123;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Brain;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udell T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalek U., Smallos D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Brain;
RA Strausberg R.;
RL Submitted (JAN-2004) to the EMBL/GenBank/DBJ databases.
EMBL; BC065123; AAH65123.1; -
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR001487; Bromodomain.
DR InterPro; IPR011011; FYVE_PHD_Znf.
DR Pfam; PF00439; Bromodomain; 1.
DR Pfam; PF00628; PHD; 1.
DR PRINTS; PR00503; BROMODOMAIN.
DR SMART; SM00297; BROMO; 1.
DR SMART; SM00249; PHD; 1.
DR PROSITE; PS00633; BROMODOMAIN_1; 1.

```
DR PROSITE; PSS0014; BROMODOMAIN_2; 1.
DR PROSITE; PSS01359; ZF_PHD_1; 1.
DR PROSITE; PSS0016; ZF_PHD_2; 1.
FT NON_TER 1
SQ SEQUENCE 892 AA; 101830 MW; 95A6C798F37AA93F CRC64;

Query Match
Best Local Similarity 24.5%; Pred. No. 1.6e-18; Indels 263; Gaps 28;
Matches 216; Conservative 137; Mismatches 264;

QY 777 SAEWLKERLAVLKEENDKKRAKQKQKKEAKNKE--NGKVENGLGKTDKRRIVK---F 831
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 4 AARIRRRKEEKLKQEQKMKQKQKLEKDEQKNSAAVPGEEEREDFDISTENKNIQKDL 63

QY 832 EPOVTEADMTISAVK-SRRLLAIQAKEREIOEREMKVKLRQA-----EERIRKHAA 886
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 64 DPDVVTEDEDDPGSHKSRGRGVGTAVKQCIQKQEBMNYCIQKQPLSADAEEALRQEQOQ 123

QY 887 AEKATQEGIAKAKLVNRRTPICTDNRHNYWLFSDVPGLFTEK---GWVHDSI-----D 938
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 124 KEKELLDKIQSAIACTNIFPLGRDLRYRYWIF-PSIPGLFTIEDYSGLTEDMLLRPSS 182

QY 939 YRFNHCKDHTVSGDEDYCPKSKANLGNASMTQHGTATEVAVETTPKQGNLWFLC 998
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 183 FHNNAAQPRDPQVS-----IKTEESFLSESTSLDQ--GPFDDSVLLPKPVHKPNRWCFY 234

QY 999 DSQKELDELLNCLHPQIGRESOLKREKRYODII-----HSIHLARKPNLGLK--- 1047
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 235 SSCAQLDQLIDALNSGRHRESALKETLLQEKSRICAQLAHFSEKRFHFDKPDQADSKPVS 294

QY 1048 -----SCDGNQ-----ELLNPLASDLIEVATRLQKGLGY----- 1077
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 295 SGRSSGACDISQMSAEROLELRDLFDLDIEDIRYQGLGAIKVTDQVWRSALENGRY 354

QY 1078 -----VEETSFEARVISLEKL---KQFCECVIALQAS----- 1107
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 355 ELLSEESKENGVIKTVNEDVEEMEQARVIVRDLGLGKTKTPTSTISASTPOSVSNV 414

QY 1108 -----VIKKFLOQFMAPKQKRKLQSEDSAKTEVDEEKQVWEAKVASAL 1153
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 415 VHYLALALFQIEQGIERRFLK--AP-----LDGNDGSRYSK-----TVL 451

QY 1154 EWMKTAIRAEOQFSRMVHLGLMDACIKWDMSEAENARCKVCPKGGEDDKLILCDRCNKAF 1213
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 452 DRWRESLLSSASLSQVFLHLSTLDRSVMMWSKILNARCKICRKGDAENNVLCDCGDRGH 511

QY 1214 HLFCLRPALYEVDPGEWQCPAQCP--ATARRNSRGRNYTE--ESASEDSEDDSEDEEE 1269
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 512 HTYCVRPKLKAVPDGDMFCPCRPKQSRRLSSRRQPSLESDEEMEEGMEDDDDDEVD 571

QY 1270 EEEEEEEDYVA-----GLRLRPRTIRGKHSVTPPAARSGR-----RP 1309
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 572 EERGSEEEYEQDEEDSDDEALSPPK--RGRPQVRLFIKTGRFGSPFSRSRQDP 629

QY 1310 GKPKHSTRSQPK-----APPVD-----DAEVDDELV 1335
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 630 GRYPERSQOSTPNTAKSASKNLRTSRAPPTETRLRVGSRSTRHSPSALQDVVFELLS 689

QY 1336 -----LQTKRSSR-----ROS 1346
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 690 PHSKRGKRGADHTPEHSPSFTNFRVSTSRSRQLIPLNTAESLSLQHSSEKRGKRS 749

QY 1347 L-----ELQKCEEILHKIVKRFSPFPVTRDAEDYDVI 1387
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 750 TESSPVPLNRRSSRGQGVHLSAFLQVLVELVRHDDSWPFLKLSKIQVPDYDIKKP 809

QY 1388 MDPQTVQNKSCSGSVRSVQEFLTDMKQVETNAEVNCRGS 1427
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 810 IALNTIIREKVNCKEYLASEFIDDIELMFNSCFEYNPRNT 849

RESULT 15
Q6YI94
```

```
ID Q6YI94 PRELIMINARY; PRT; 1698 AA.
AC Q6YI94;
DT 05-JUL-2004 (TRENBLrel. 27, Created)
DT 05-JUL-2004 (TRENBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TRENBLrel. 27, Last annotation update)
DE Putative chromatin remodelling factor BAZ2A.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
OC xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1] SEQUENCES FROM N.A.
RP Ruzov A., Meehan R.;
RA Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
RL EMBL; AY145834; AAN61105.1; -.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR000637; A-T hook.
DR InterPro; IPR001487; Bromodomain.
DR InterPro; IPR004022; DDT.
DR InterPro; IPR001739; Methyl-CpG_bind.
DR InterPro; IPR001965; Znf_PHD.
DR Pfam; PF02178; AT_hook; 3.
DR Pfam; PF00439; Bromodomain; 1.
DR Pfam; PF02791; DDT; 1.
DR Pfam; PF01429; MBD; 1.
DR Pfam; PF00628; PHD; 1.
DR PRINTS; PR00503; BROMODOMAIN.
DR SMART; SM00384; AT_hook; 3.
DR SMART; SM00297; BROMO; 1.
DR SMART; SM00571; DDT; 1.
DR SMART; SM00391; MBD; 1.
DR SMART; SM00249; PHD; 1.
DR PROSITE; PS00014; BROMODOMAIN_2; 1.
DR PROSITE; PS00827; DDT; 1.
DR PROSITE; PS00016; ZF_PHD_2; 1.
DR SEQUENCE 1698 AA; 192135 MW; F62D613AD9611B6F CRC64;

Query Match
Best Local Similarity 6.9%; Score 550; DB 2; Length 1698;
Matches 307; Conservative 210; Mismatches 492; Indels 576; Gaps 58;

QY 143 VDEATEKKSD---GACDSPSDKENSQ-----IAQHQKQKTVVDEGR 186
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 353 MQESAEMGEDPEGSKABEPVSGPENVSQDEMTIENTSPGPCAAHEEELEPEGEVK- 411

QY 187 RESINDRARRSPKL-----PTSLKKGKRWAPPKPLPHKYDVVKLQWEDKIISNV 236
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 412 --TISRRRIATPEQVCFPLQHGWRREVRIKKGSHRWQGETWTYYAPCGKMKQFPFV 469

QY 237 P--ADSLRTER--PPNKEIVRYEIRHNALRAAGTGENAPWYVEDELVKYSLPSK 291
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 470 SKNAGPVFRREHFSFSPMPVGDVFEERN---FAEGLQWV---LLNSEIIPSR 517

QY 292 LLDPVYKTYMLNPSTKRVKTSKSKTNSLSPLNPKLWCHVHLKSLKSGSPL 351
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 518 -----IAITGKR---GRPNLEKAKAEQKA-----KRGGRPP 548

QY 352 KYKNSKNSKSPHEHLEENMKMSPNKLHNTFHI PKKGPAPKPKGKSKDKPLAKGRSKGI 411
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 549 KVK-----MIDLLS-----KADAKLRKLENQDI 572

QY 412 LMGQKSTGNSKSPKGLKTKTKMQMTLLDMAKGTQKMTAPRNSGGTPTRTSSKPHKL 471
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 573 L-----SDSEK-----VOLCKLKKMRRKARQ----- 595

QY 472 PPAALHLIAYKENKDRKRSALSCVISTARLLSSEDRALPELRLVOKRYELLEH 531
D  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 596 -BAKLEAAAKLKEIKEKEKKQKI-----QAKNQEKAKN-----QE 631

QY 532 KKWASMSSEQRKEYLKKKREEL--KKKLKAKERKERREKLEKQRYEDQELTG-K 588
```

[illegible]

1581 RMGTRSQSPDLTFCFIIIMLESHEDAWPFLEPNPRLVGYRKTIKNPMDFMSTRHKL 1640

1399 CGSYRSVQEEFLTDWKQVFTNAEYVN 1423

1641 NGVYSRCEEFAEDALIFSNQCLFN 1665

RESULT 16

BA2B HUMAN STANDARD; PRT; 1972 AA.

AC Q9UIF8, Q96EA1; Q96SQ8; Q9P252; Q9Y4N8; 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 05-JUL-2004 (Rel. 44, Last annotation update)

DE Bromodomain adjacent to zinc finger domain 2B (hWALp4).

GN Name=BAZ2B; Synonyms=KIAA1476;

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

NCBI_TaxID=9606;

[1]

RN SEQUENCE FROM N.A. (ISOFORM 1).

RC TISSUE=Testis;

RX MEDLINE=20130112; PubMed=10662543; DOI=10.1006/geno.1999.6071; Jones M.H., Hamana N., Nezu J., Shimane M.;

RA "A novel family of bromodomain genes.;"

RL Genomics 63:40-45(2000).

[2]

RN SEQUENCE OF 753-1972 FROM N.A.

RP OHara O., Nagase T., Kikuno R.;

RA Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.

[3]

RN SEQUENCE OF 753-1972 FROM N.A.

RP MEDLINE=20273482; PubMed=10819331;

RX Nagase T., Kikuno R., Ishikawa K.-I., Hiroseawa M., Ohara O.;

RA "Prediction of the coding sequences of unidentified human genes. XVII. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro.;"

RL DNA Res. 7:143-150(2000).

[4]

RN SEQUENCE OF 1-776 FROM N.A. (ISOFORM 1).

RC TISSUE=Melanoma;

RA Ansoerge W., Wlkrner U., Mewes H.-W., Weil B., Wiemann S.;

RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.

[5]

RN SEQUENCE OF 1-799 FROM N.A. (ISOFORM 2).

RC TISSUE=Skeletal muscle;

RX MEDLINE=323388257; PubMed=12477932; DOI=10.1073/pnas.242603899; Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Klausner R.D., Collins F.S., Wagner L., Shenmen C.F., Schuler G.D., Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.M., Bhat N.K., Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heieh F., Diatchenko L., Marusina K., Farmer A.C., Rubin G.M., Hong L., Chappleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E., Brownstein M.J., Utsid T.B., Toshiyuki S., Carninci P., Prange C., Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J., Bosak S.A., McSwan P.J., McKernan K.J., Walek J.A., Gunaratne P.H., Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W., Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Fahey J., Hailton E., Kettelman M., Madan A., Rodrigues S., Sanchez A., Whiting J., Maitan A., Young A.C., Shevchenko Y., Bouffard G.G., Blakeley R.W., Townhond J.W., Green E.D., Dickson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E., Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;

RT "Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences.;"

RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

[6]

RN SEQUENCE OF 427-1215 FROM N.A. (ISOFORM 3).

RX PubMed=14702039; DOI=10.1038/ng1289; Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R., Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,

RA Sekine M., Obayashi M., Nishi T., Shibahara T., Tanaka T., Ishii S.,
RA Yamamoto J.-I., Saio K., Kawai Y., Isono Y., Nakanura Y.,
RA Nagahari K., Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M.,
RA Shiratori A., Sudo H., Hosoi T., Kaku Y., Kodaira H., Kondo H.,
RA Sugawara M., Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E.,
RA Omura Y., Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa M.,
RA Yanazaki M., Ninomiya K., Ishibaishi T., Yamashita H., Murakawa K.,
RA Fujimori K., Tanai H., Kimata M., Watanabe M., Hiraoka S., Chiba Y.,
RA Ishida S., Ono Y., Takiguchi S., Watanabe T., Yosida M., Horuta T.,
RA Kusano J., Kanehori K., Takahashi-Fujii A., Hara H., Tanase T.-O.,
RA Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K., Arita M.,
RA Imose N., Mueashino K., Yuuki H., Oshima A., Sasaki N., Aotsuka S.,
RA Yoshikawa Y., Matsunawa H., Ichihara T., Shiohara N., Sano S.,
RA Moriya S., Momiya H., Satoh N., Takami S., Terashima Y., Suzuki O.,
RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,
RA Higashigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,
RA Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,
RA Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiwara T.,
RA Ono T., Yamada K., Fujii Y., Ozaki K., Hirao M., Ohmori Y.,
RA Kawabata A., Hikiji T., Kobatake N., Ingaki H., Ikema Y., Okamoto S.,
RA Okitani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,
RA Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,
RA Toqashi T., Oyama M., Hata H., Watanabe M., Komatsu T.,
RA Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,
RA Okumura K., Nagase T., Nomura N., Kikuchi H., Masuho Y., Yamashita R.,
RA Nakai K., Yada T., Nakamura Y., Ohara O., Isegai T., Sugano S.,
RT "Complete sequencing and characterization of 21,243 full-length human
RT cDNAs.";
RL Nat. Genet. 36:40-45 (2004).
RN [7].
RP SEQUENCE OF 1524-1972 FROM N.A.
RC TISSUE-Testis;
RA Poustka A., Klein M., Mewes H.-W., Gassenhuber J., Wiemann S.;
RL Submitted (JUN-1999) to the EMBL/GenBank/DBSJ databases.
CC -!- FUNCTION: May play a role in transcriptional regulation
CC interacting with ISWI.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=3;
CC Comment=Experimental confirmation may be lacking for some
CC isoforms;
CC Name=1;
CC IsoId=Q9UIF8-1; Sequence=Displayed;
CC Name=2;
CC IsoId=Q9UIF8-2; Sequence=VSP_000553;
CC Note=Inferred from Ref.4;
CC Name=3;
CC IsoId=Q9UIF8-3; Sequence=VSP_000554;
CC Note=Inferred from Ref.5;
CC -!- TISSUE SPECIFICITY: Expressed at varying levels in several
CC tissues, whereas a smaller transcript was expressed specifically
CC in testis.
CC -!- SIMILARITY: Belongs to the WAL family.
CC -!- SIMILARITY: Contains 1 bromodomain.
CC -!- SIMILARITY: Contains 1 DDT domain.
CC -!- SIMILARITY: Contains 1 methyl-CpG-binding (MBD) domain.
CC -!- SIMILARITY: Contains 1 PHD-type zinc finger.
CC -!- CAUTION: Ref.5 sequence differs from that shown due to a
CC frameshift in position 731.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AB032255; BAA89212.1; --
CC EMBL; AB040909; BAA96000.2; ALT INIT.
CC EMBL; AL834381; CAD39044.1; ALT INIT.
CC EMBL; BC012576; AAH12576.1; ALT FRAME.
CC EMBL; AK027612; BAB55231.1; ALT_INIT.
CC EMBL; AL080173; CAB45759.1; --

DR PIR; T12495; T12495.
DR HSSP; Q9UIG0; 1F62.
DR Genew; HGNC:963; BAZ2B.
DR MIM; 605683; --
DR InterPro; IPR001487; Bromodomain.
DR InterPro; IPR004022; DDT.
DR InterPro; IPR001739; Methyl-CpG_bind.
DR InterPro; IPR001965; Znf_PHD.
DR Pfam; PF00439; Bromodomain; 1.
DR Pfam; PF02791; DDT; 1.
DR Pfam; PF01429; MBD; 1.
DR Pfam; PF00628; PHD; 1.
DR PRINTS; PS00503; BROMODOMAIN.
DR PROSITE; PS00633; BROMODOMAIN_1; FALSE_NEG.
DR PROSITE; PS00104; BROMODOMAIN_2; 1.
DR PROSITE; PS50827; DDT; 1.
DR PROSITE; PS50982; MBD; 1.
DR PROSITE; PS01359; ZF_PHD_1; FALSE_NEG.
DR PROSITE; PS50016; ZF_PHD_2; 1.
KW Alternative splicing; Bromodomain; Coiled coil; DNA-binding;
KW Nuclear protein; Transcription regulation; Zinc-finger.
FT DOMAIN 543 614 MBD.
FT ZN_FING 1735 1785 PHD-type.
FT DOMAIN 1881 1951 Bromodomain.
FT DOMAIN 45 69 Ser-rich.
FT DOMAIN 73 79 Poly-Glu.
FT DOMAIN 399 470 Asp/Glu-rich (acidic).
FT DOMAIN 641 672 Arg-rich.
FT DOMAIN 687 865 Coiled coil (Potential).
FT DOMAIN 706 875 Lys-rich.
FT DOMAIN 1100 1143 Asp-rich.
FT DOMAIN 1138 1179 Coiled coil (Potential).
FT VARSP_LIC 437 534 Missing (in isoform 2).
FT VARSP_LIC 593 626 Missing (in isoform 3).
FT CONFLICT 137 137 S -> F (in Ref. 1).
FT CONFLICT 226 226 S -> L (in Ref. 1).
FT CONFLICT 627 627 G -> E (in Ref. 1).
FT CONFLICT 722 722 E -> K (in Ref. 5).
FT CONFLICT 790 799 RIKEKERRQ -> QKKKKKKKK (in Ref. 5).
FT CONFLICT 1128 1128 K -> Q (in Ref. 1).
FT CONFLICT 1183 1183 Q -> P (in Ref. 1).
FT CONFLICT 1195 1195 Q -> R (in Ref. 1).
FT CONFLICT 1453 1453 L -> S (in Ref. 3).
FT CONFLICT 1838 1838 K -> Q (in Ref. 1).
SQ SEQUENCE 1972 AA; 220709 MW; D4F105C927B91569 CRC64;
Query Match 6.7%; Score 537; DB 1; Length 1972;
Best Local Similarity 18.0%; Pred. No. 7.2e-14; Indels 758; Gaps 68;
Matches 354; Conservative 258; Mismatches 595;
QY 8 KPFLVNLPLGPEEPF-FTIPH-----TQFAFRTRREYERLERYSERIWTCKST 55
DB 189 KPLSLVNOAKKETYNKLVPSDPVLKAGNKTSESSLSLTSLSKREY-----KQA 241
QY 56 GSSQLTHKEAMEEEOEVAELK-----EEFPAYEKLVLEVMVHN-----TA 97
DB 242 FPSQLKQESSKSLKVAALSNPKATSSSPA-HPKQTLNHNPNPFLTNALLGNHPNG 300
QY 98 SLEKLVDTAWLEIMTKYAVGECDFEV-----124
DB 301 VIQSVIQAIPALATTKTKQSKINENIAAASPTFPSPVNLSTSGRRTPGNTFPVNPAS 360
QY 125 -----GKEKML--KVKIVKI-----HPL-----EKVDEE 146
DB 361 PILHSQGEKAVNNVNPVKQHHSHPAKSLVEQFRGTDSDIPSSKSDSDEDEEDDE 420
QY 147 ATEKSKDGACDPSDDKNSGSOIADHOKETVVEDEGRRSINDRARRSPKRLTSLK 206
DB 421 EDEDEDDDDSDSQSSDSNSDTEGSEEDDDKQDQSDSD-----T 467

Qy	207	KG	ER	KA	WP	PK	FL	PH	KY	DI	VK	LQ	NE	KI	IS	NV	---	PA	DS	LI	TR	EP	PN	KE	I	VR	YF	IR	HA	LA	264	
		:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	
Db	468	E	G	E	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	503	
Qy	265	GT	GE	NA	PV	VE	DE	LV	KY	SL	PS	KS	DF	LL	DP	KY	MT	NP	ST	KT	KN	TS	GP	DR	KS	PK	SK	K	T	323		
Db	504	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---		
Qy	324	DN	SS	LP	NP	KJ	CH	VH	LK	SL	GS	SP	LK	VN	KN	SK	SP	---	---	---	---	---	---	---	---	---	---	---	---	376		
Db	545	DE	RE	L	P	LE	Y	G	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	600		
Qy	377	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	395		
Db	601	DI	SR	DN	FS	SA	KI	RV	GD	F	Y	EA	RD	GP	Q	GM	OW	CL	KE	ED	VI	PR	AM	EG	RR	GP	PN	DR	QA	660		
Qy	396	KH	SK	PL	KA	K	G	R	S	K	G	I	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---		
Db	661	RE	ES	MR	RR	K	GR	PP	V	N	G	NA	EF	LD	NA	DA	K	L	R	K	L	QA	ET	AR	QA	Q	I	K	L	R	720	
Qy	422	K	S	P	K	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---		
Db	721	KE	AK	Q	QA	I	MA	EE	B	K	R	K	Q	E	Q	I	K	I	M	K	Q	E	K	I	R	---	---	---	---	---		
Qy	480	A	Y	I	EN	K	D	R	E	K	S	A	L	S	C	V	I	S	T	A	R	L	S	S	E	D	A	R	L	P	E	539
Db	770	E	A	K	K	K	E	E	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---		
Qy	540	BE	OR	KE	Y	L	K	K	E	E	K	K	E	K	E	R	E	K	E	M	L	E	K	R	E	D	E	---	---	---	---	
Db	810	ER	R	Q	H	M	L	M	K	A	E	R	K	A	E	B	E	R	L	K	O	B	R	D	E	K	L	N	K	E	R	869
Qy	591	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---		
Db	870	M	C	L	A	D	Q	X	P	L	P	R	I	P	G	L	V	S	G	S	T	F	S	D	C	L	V	V	Q	F	L	927
Qy	643	A	D	K	G	F	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---		
Db	928	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---		
Qy	696	V	R	L	C	R	S	D	V	Q	E	S	E	G	S	D	T	D	N	K	S	A	F	E	D	N	E	V	O	F	E	755
Db	984	L	Q	I	P	M	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---		
Qy	756	L	C	H	R	I	M	T	S	V	Q	D	H	M	T	R	Q	M	S	A	E	L	K</									

RESULT 17

Q68DI8 PRELIMINARY; PRT; 1905 AA.
Q68DI8
ID Q68DI8
AC Q68DI8;
DT 25-OCT-2004 (TtEMBLrel. 28, Created)
DT 25-OCT-2004 (TtEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TtEMBLrel. 28, Last annotation update)
DE Hypochemical protein DKF2p781B109.
GN Name=DKF2p781B109;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Euteleostomi;
OC NCBI_TaxID=9606;
OX [1]
RN SEQUENCE FROM N.A.
RP TISSUE=Fetal kidney;
RC The German cDNA Consortium;
RG Bloecker H., Boecker M., Brandt P., Mewes H.W., Weil B., Amid C.,
RA Osanger A., Fobo G., Han M., Wiemann S.;
RL Submitted (AUG-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; CR749379; CAH18232.1; -;
DR InterPro; IPR000637; A+T hook.
DR InterPro; IPR001487; Bromodomain.
DR InterPro; IPR004022; DDT.
DR InterPro; IPR011011; FYVE PHD ZnF.
DR InterPro; IPR001739; Methyl-CpG_bind.
DR InterPro; IPR001965; ZnF_PHD.
DR Pfam; PF02178; AT hook; 4.
DR Pfam; PF00439; Bromodomain; 1.
DR Pfam; PF02791; DDT; 1.
DR Pfam; PF01429; MBD; 1.
DR Pfam; PF00628; PHD; 1.
DR PRINTS; PR00503; BROMODOMAIN.

RP SEQUENCE OF 1038-1878 FROM N.A.
RC TISSUE=lymph;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Dege J.G.,
RA Klausner R.D., Collins P.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Munzy D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -!- FUNCTION: May play a role in transcriptional regulation
interacting with ISWI. May serve a specific role in maintaining or
altering the chromatin structure of the rDNA locus (By
similarity).
CC -!- SUBUNIT: Together with ISWI/SNF2h, it forms a complex termed NoRC
(nucleolar remodeling complex).
CC -!- SUBCELLULAR LOCATION: Nuclear. Colocalizes with the basal RNA
polymerase I transcription factor UBF in the nucleolus.
CC -!- TISSUE SPECIFICITY: Expressed at moderate levels in most tissues
analyzed, including heart, brain, placenta, lung, skeletal muscle,
kidney and pancreas.
CC -!- SIMILARITY: Belongs to the WAL family.
CC -!- SIMILARITY: Contains 4 A.T hook DNA-binding repeats.
CC -!- SIMILARITY: Contains 1 bromodomain.
CC -!- SIMILARITY: Contains 1 DDT domain.
CC -!- SIMILARITY: Contains 1 methyl-CpG-binding (MBD) domain.
CC -!- SIMILARITY: Contains 1 PHD-type zinc finger.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL; AB032254; BAA89211.1; -;
CC EMBL; AF000422; AAB60864.1; -;
CC EMBL; AB002312; BAA20773.1; -;
CC EMBL; BC008965; AAH08965.2; -;
CC HSSP; Q9UIG0; 1F62.
CC Genew; HGNC:962; BAZ2A.
CC MIM; 605682; -;
CC GO; GO:0005731; C:nucleolus organizer complex; NAS.
CC GO; GO:0003677; F:DNA binding; NAS.
CC GO; GO:0030528; F:transcription regulator activity; NAS.
CC GO; GO:0006338; P:chromatin remodeling; NAS.
CC GO; GO:0006355; P:regulation of transcription, DNA-dependent; NAS.
CC InterPro; IPR000637; A+T hook.
CC InterPro; IPR001487; Bromodomain.
CC InterPro; IPR004022; DDT.
CC InterPro; IPR001739; Methyl-CpG_bind.
CC InterPro; IPR001965; Znf_PHD.
CC Pfam; PF02178; AT hook; 4.
CC Pfam; PF00439; Bromodomain; 1.
CC Pfam; PF02791; DDT; 1.
CC Pfam; PF01429; MBD; 1.
CC Pfam; PF00628; PHD; 1.
CC PRINTS; PF00503; BROMODOMAIN.
CC SMART; SM00384; AT_hook; 4.

DR SMART; SM00297; BROMO; 1.
DR SMART; SM00571; DDT; 1.
DR SMART; SM00391; MBD; 1.
DR SMART; SM00249; PHD; 1.
DR PROSITE; PS00633; BROMODOMAIN_1; FALSE_NEG.
DR PROSITE; PS50014; BROMODOMAIN_2; 1.
DR PROSITE; PS50827; DDT; 1.
DR PROSITE; PS50982; MBD; 1.
DR PROSITE; PS01359; ZF_PHD_2; 1.
DR PROSITE; PS01359; ZF_PHD_2; 1.
KW Bromodomain; Coiled coil; DNA-binding; Nuclear protein; Repeat;
KW Transcription regulation; Zinc-finger.
FT DOMAIN 519 590 MBD.
FT DNA_BIND 622 634 A.T hook 1.
FT DNA_BIND 643 655 A.T hook 2.
FT DOMAIN 821 886 DDT.
FT DNA_BIND 1159 1171 A.T hook 3.
FT DNA_BIND 1377 1389 A.T hook 4.
FT ZN_FING 1649 1699 PHD-type.
FT DOMAIN 1783 1853 Bromodomain.
FT DOMAIN 633 772 Lys-rich.
FT DOMAIN 666 765 Coiled coil (Potential).
FT DOMAIN 1185 1250 Glu-rich.
FT DOMAIN 1283 1384 Pro-rich.
FT DOMAIN 1732 1735 Poly-Arg.
FT CONFLICT 574 574 L -> V (in Ref. 2).
FT CONFLICT 700 700 L -> Q (in Ref. 2 and 3).
FT CONFLICT 720 720 H -> Q (in Ref. 2 and 3).
FT CONFLICT 727 738 K -> R (in Ref. 3).
FT CONFLICT 785 785 P -> L (in Ref. 3).
FT CONFLICT 951 951 GR -> EG (in Ref. 3).
FT CONFLICT 1005 1006 G -> S (in Ref. 1).
FT CONFLICT 1163 1163 R -> L (in Ref. 1).
FT CONFLICT 1166 1166 S -> P (in Ref. 1).
FT CONFLICT 1172 1172 L -> F (in Ref. 1).
FT CONFLICT 1178 1178 A -> V (in Ref. 1).
FT CONFLICT 1202 1202 P -> L (in Ref. 1).
FT CONFLICT 1292 1292 L -> F (in Ref. 1).
FT CONFLICT 1295 1295 P -> L (in Ref. 1).
FT CONFLICT 1313 1313 Missing (in Ref. 4).
FT CONFLICT 1407 1410 R -> P (in Ref. 1).
FT CONFLICT 1416 1416 R -> P (in Ref. 1).
FT CONFLICT 1541 1541 E -> K (in Ref. 1).
FT CONFLICT 1571 1571 E -> I (in Ref. 1).
FT CONFLICT 1616 1616 B -> Q (in Ref. 1).
FT CONFLICT 1622 1622 Q -> H (in Ref. 1).
FT CONFLICT 1629 1629 Q -> H (in Ref. 1).
FT CONFLICT 1636 1636 R -> K (in Ref. 1).
FT CONFLICT 1739 1739 G -> R (in Ref. 1).
FT CONFLICT 1754 1754 G -> R (in Ref. 1).
SQ SEQUENCE 1878 AA; 208480 MW; 400970CA68234317 CRC64;
Query Match 6.6%; Score 525.5; DB 1; Length 1878;
Best Local Similarity 19.3%; Pred. No. 2e-13;
Matches 263; Conservative 176; Mismatches 446; Indels 479; Gaps 40;
QY 415 OKSTGNSKSPKGLKTPKTKOMTLLDMAKGTQ---KMTAPRNSGGTPRTSSPKPHKL 471
Db 619 QAITGRGRPR---NTEAKTKEVKVGRGRPPKVKITEL-----LNKTDNRPLKKL 669
QY 472 PPAALHLIAYYKENKDRDKRSALSCVISTKARLLSSDRARLPBELRSLVQRYELLEPH 531
Db 670 EA-----QETLNEDKAKIAKSKKMRQVQVGECLTTIQGARNKRQETSLKH 720
QY 532 KRWASMBEQRKEVYLKKRBEELKKLKEKAKER---REKE----- 569
Db 721 KE--AKGKSKAEKKGKTKQEKLEKVKREKVKMKREKVEVTKAKPACKADKTALATOR 778
QY 570 MLERLEKO-----KRYEDQELTG-KNLPAPFLVDTPG--LPLTFLGDVAMVVEPLS 618
Db 779 RLREERQKQOMILEEMKKPTEDMCLTDHQLPDPFSV---PGULFSGAFSDCLTIVEFLH 835
QY 619 CYSGLLLPDAQYPITAVS-LMEALSADKGGFLYLNRLVILLQTLQTLQLODEIADYGE 677

Db 836 SFGKVLGDPKADVPSGLVQLGSLGCGSLGEVDLLVRLKKAALHDPGRPSYCSQSLKI 895
QY 678 LGMKLSLPLTLHSLVSLVRLCLRSDDVOEESGSDTDNDKDSAAPENEVEQDFLEKLE 737
Db 896 LGEKVSLEPLTDNVSLEILRCFLMAYG-----VEPALCDRLR 932
QY 738 TSEFFELTSEELQLILTALCHRI-----LMTYSVQDHMETRQMSAELM--KERLAVLEE 791
Db 933 TOPFOAQQPQAAVLAPVPHLNGSTLIINEIDKTLSEMSYRKNKMIVEGRLLRK-- 990
QY 792 NDKGAEKOKRMEAKNKGKGVNGKTKRKRIKVFEPQVQVTEAEDMISAVKSRRL 851
Db 991 --TVLAKRTGRSEVM-----GRPECLGRSSRIMEETSCWBEESIESIAAPGRR- 1042
QY 852 LAIOAKKEREIOEREMKV-KLBRQAEERIRKHAAAEKAFQEGIAKAKLVNRRTPIGTD 910
Db 1043 ----GRDGEVDATASSIPELEROIEKLSK-----QLFFRKGLHSSQMLRAVSLGQD 1092
QY 911 RNHNRYWLPDSVPGLFIE-----KGWVHDSIDYRF----- 941
Db 1093 RYRRRIWVL-PYLAGIFVGTGSLVPEVVIKKTDSLKVAHAHSLPALFSKMKELAGS 1151
QY 942 -----NHHCKDHTYSGEDYCPRSK----- 961
Db 1152 NTTASSPARARGPRKTPGSMQPRHLKS-PVRGQDSEQPAQLQPEAQLHAPAQPOQL 1210
QY 962 -----KANLGN-----ASMNTQGTATEVAVETTTKQG-- 991
Db 1211 QLQLQSHKGFLREGSPLSGQSDLSQSAFLSLSQTSQSHSLSSSVLTPDSSPGKL 1270
QY 992 -----QNLWFLCDSO----- 1001
Db 1271 DPAPSQPPEPDEAESPDQALWFNISAQMPNAAPTPPAVSEDOPTSPQOLASS 1330
QY 1002 ----- 1001
Db 1331 KPMNPSAANPCSPVQFSSTPLAGLAPKERRAGDPCGEMPOSPTGLQPKRGRPPSKFKQ 1390
QY 1002 -----KELDELNCLHPQIGIRSOLKLEKRYQDII 1033
Db 1391 MEQRYLTQLTAQVPVPEMCGWWIRDPMLDAMLKALHPRGIRGKALHKLNK-HRDFL 1449
QY 1034 HSIHL--ARKPNLIGKSCDGNQD--LNLPLRSDLIEVATLQGGIGYVEETSEFARVIS 1090
Db 1450 QEVCLRPSADPIFEPRQLPAPQEGIMSWSPKE-----KTYETDLAVLQWVLELSQRIJM 1503
QY 1091 -----LEKLKDFGECV----- 1101
Db 1504 SDLQIRGWTCPSPDSTREDLAYCEHLSDSQEDITWRGRGREGLAQPKTTNPLDLAVMRL 1563
QY 1102 IALQASVIKFLGFMAPKQKRKLQSDSAXTEEVDEEKKVVE-----EAK 1148
Db 1564 AALEQNVERRYLRPLWP-----THEVYLEKALLSTPNGAPEGTTTTEISYE 1609
QY 1149 VASALEKWKTAIREAQTFSRMHVLLGMLDACIKWDMASNAENARCKVCPKKGEDDKLILCDE 1208
Db 1610 ITPRIRVWQTLERCKSAQAQCLGQLERSIANEKSVMKVTCLVCRKGDNDFFLLCDG 1669
QY 1209 CNKAHFLCLRPALEVPDGEQCPACQAPATARRNSRGRNYTEESASESDESDSEDEE 1268
Db 1670 CDRGCHYIHRPMEAVPEGDFNFCIVC-----LAQQVEGEFTQKPGF 1711
QY 1269 EEEEEEEEDYEV--AGURLRPRKTIIRKSHSVIPPAARSGRRPGKKPHSTRSQPKAPP 1325
Db 1712 PKRGQRKSGYSLNFSEGDGRRRRVLLRGRES--PAAQPRYSEEGSLSPSKRRR----- 1762
QY 1326 VDAAEVDVLQTKSSRRQSLELOKCEIILHKIVKYPFSWPPREPVTRDEAEDYDVIT 1385
Db 1763 -----LSMRNHHSDLTFCEIILMEMESHDAWFFLEFPVNPRLVSGYTRRIK 1808
QY 1386 HPMDFQTVQKSCGSYSRVQBFLLTDMKQVFTNAEYVNCRGSHV 1429

Db 1809 NPMDFSTMRLRLRGYTSSEFAADALLVFDNQCOTFNEDDSEV 1852
RESULT 19
B2A2A_MOUSE
ID B2A2A_MOUSE STANDARD; PRT; 1850 AA.
AC Q91VE5;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Bromodomain adjacent to zinc finger domain 2A (Transcription
termination factor-1 interacting protein 5) (TTF-1 interacting protein
5) (Tip5).
GN Name=Baz2a; Synonyms=Tip5;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
SEQUENCE FROM N.A.
MEDLINE=21423468; PubMed=11532953; DOI=10.1093/emboj/20.17.4892;
RA Strohnner R., Nemeth A., Jansa P., Hoffmann-Rohrer U., Santoro R.,
RA Laengst G., Grummt I.;
RT "NORC-a novel member of mammalian ISWI-containing chromatin remodeling
machines.";
RL EMBO J. 20:4892-4900 (2001).
CC -!- FUNCTION: May play a role in transcriptional regulation
interacting with ISWI. May serve a specific role in maintaining or
altering the chromatin structure of the rDNA locus.
CC -!- SUBUNIT: Together with ISWI/SNF2h, it forms a complex termed NORC
(nucleolar remodeling complex).
CC -!- SUBCELLULAR LOCATION: Nuclear. Colocalizes with the basal RNA
polymerase I transcription factor UBF in the nucleolus.
CC -!- SIMILARITY: Belongs to the WAL family.
CC -!- SIMILARITY: Contains 4 A.T hook DNA-binding repeats.
CC -!- SIMILARITY: Contains 1 bromodomain.
CC -!- SIMILARITY: Contains 1 DDT domain.
CC -!- SIMILARITY: Contains 1 methyl-CpG-binding (MBD) domain.
CC -!- SIMILARITY: Contains 1 PHD-type zinc finger.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

CC EMBL; AJ309544; CAC69992.1; -.
CC HSSP; Q9UIG0; 1F62.
CC MGD; MGI:2151152; Baz2a.
CC GO; GO:0005731; C:nucleolus organizer complex; ISS.
CC GO; GO:0003677; F:DNA binding; ISS.
CC GO; GO:0030528; F:transcription regulator activity; ISS.
CC GO; GO:0006338; P:chromatin remodeling; ISS.
CC GO; GO:0006355; P:regulation of transcription, DNA-dependent; ISS.
CC InterPro; IPR000637; A-T hook.
CC InterPro; IPR001487; Bromodomain.
CC InterPro; IPR004022; DDT.
CC InterPro; IPR001739; Methyl-CpG_bind.
CC InterPro; IPR001965; Znf PHD.
CC Pfam; PF02178; AT hook; 4.
CC Pfam; PF00439; Bromodomain; 1.
CC Pfam; PF02791; DDT; 1.
CC Pfam; PF01429; MBD; 1.
CC Pfam; PF00628; PHD; 1.
CC PRINTS; PR00503; BROMODOMAIN.
CC SMART; SM00384; AT hook; 4.
CC SMART; SM00297; BROMO; 1.
CC SMART; SM00571; DDT; 1.
CC SMART; SM00391; MBD; 1.
CC SMART; SM00249; PHD; 1.
CC PROSITE; PS00633; BROMODOMAIN_1; FALSE_NEG.

DR PROSITE; PS50014; BROMODOMAIN_2; 1.
 DR PROSITE; PS50827; DDT; 1.
 DR PROSITE; PS50982; MBD; 1.
 DR PROSITE; PS01359; ZF_PHD_1; FALSE_NEG.
 DR PROSITE; PS50016; ZF_PHD_2; 1.
 KW Bromodomain; Coiled coil; DNA-binding; Nuclear protein; Repeat;
 KW Transcription regulation; Zinc-finger.
 FT DOMAIN 499 570 MBD.
 FT DNA_BIND 602 614 A.T hook 1.
 FT DNA_BIND 623 635 A.T hook 2.
 FT DOMAIN 800 865 DDT.
 FT DNA_BIND 1137 1149 A.T hook 3.
 FT DNA_BIND 1351 1363 A.T hook 4.
 FT ZN_FING 1623 1673 PHD-type.
 FT DOMAIN 1755 1825 Bromodomain.
 FT DOMAIN 613 738 Lys-rich.
 FT DOMAIN 647 774 Coiled coil (Potential).
 FT DOMAIN 1006 1012 Poly-Glu.
 SQ SEQUENCE 1850 AA; ESDD4FED1D4DAE1 CRC64;
 Query Match 6.4%; Score 512.5; DB 1; Length 1850;
 Best Local Similarity 18.9%; Pred. No. 6.8e-13;
 Matches 292; Conservative 205; Mismatches 456; Indels 593; Gaps 55;
 QY 298 YMTLNPSTKXNTGSPDRPKSKTDSLSPLNPKLWCHVHLKSLGSPKLVKNSK 357
 DB 458 FQTVSPA--RKNVSSAPKARADRETTGGAVA-----VSGSGDVLK---R 497
 QY 358 NKSPEEHLKMMKMSPKLHNFHPK-----KGPAPKPGKHSKPKLAKG 406
 DB 498 RIATPEE-----VRLPLQHWRRREVRIKKGSHWQGETWYGPCGRMKQF---PEVIKY 549
 QY 407 RSKGILNGKSTGNSKSPKGL-----KTPK-TKMKQMTLDMAGTKMTAPRNGS 458
 DB 550 LSENVVSHVREHFSFSPMPVGDFFERTDPGLQWQLSABEISPRQAIT-----GKR 605
 QY 459 GTPRTSK-PHKHLP-----PAALHLIAYYKKNDRKRSALSVCWISKRTARLSSE 509
 DB 606 GRPRNKEKANKVPRKGRGRPPKIKMPELLNKTDLNLPKK-----LETQELTSD 658
 QY 510 DRARLPEELSLVOK--RYELL-----EHKKWASMSBEORKEYLKKKEELK--- 555
 DB 659 DKAQMTKNKKMKQKQVORGESQTFVQOQARNKKQDTKSLKQKDTKKLKAERKMTQ 718
 QY 556 KKLKEKAK-ERREKEML-----ERLEKQKRYEQE----- 584
 DB 719 EKLKEKVEKKEKVKAKGEGPRAPRSCADKTLATQKRLBQQQQAILEMKKPTG 778
 QY 585 ----LTKGNLPAFLVDTPEG--LPNTLFGDVAMVVEFLSCYGLLPPDAQYPITAVS-LM 638
 DB 779 MCLSDHQPLPDFTRI---PGLTSSRAFSDCLTIVEFLHSFGKVLGFDLTQDVPSLGLVQ 835
 QY 639 EAL-----SADKGGFLYNRLVLILOTLLOLDEIAEDYCE-----LGMKLSLPLTL 689
 DB 836 EGLLCCGSLDK-----VODLLVRLKALHDPGLPPYQCQSLKILGERKMSLPLTR 886
 QY 690 HVSSELVRLCLRRSDVOEESGSDTDNDKDSAAFEDNEVDLEKLETSEFELTSEK 749
 DB 887 DNVEILRCFLMA-----YRVEPPFCDLSLRTQFQAQPPQOK 923
 QY 750 LQILTALCHRI-----LMTYSVDHMETROQMSAELW--KERLAVLKEENDKKRAEKQK 803
 DB 924 AAILAFLVHELNSSTIIINEIDKTLSSVSSCRKNKWIVEGRLLRLKLTAKRTGRPEVM 983
 QY 804 EMEAKNKGVKNGELGKTRKRIKVPFQV--DTEADMIANVKSRRLLATQAKKERI 862
 DB 984 E-----GAEEDGLGR--RSRIMEETSGIEEEEEENTTAVHGRR-----GRKEGI 1028
 QY 863 QEREMKV-KLERQAEERIRKHKAAEKAFOEGIAKAKLVMRRTPIGTDRNHNRYLFS 921
 DB 1029 DVNAASSIPLEERHIEKLSK-----QLFRKKLLHSSQMLRAVSLGQDRYRRHWVL-P 1081
 QY 922 EYVGLFIEKGWVHDSIDYRFNHHCKDHTVSGDE-----DYCPRSKKANL----- 965

DB 1082 YLAGIFEVGS-----EGSTVTEDEIKQETSLMEVVTSTPSSARASVKREL 1127
 QY 966 -GKNASMN----- 972
 DB 1128 TGSNASTSPARSRGRPRKPKPGSLQPHLOSTIRECDSSBOAQOVHPEPQLOAQTPQH 1187
 QY 973 -----TQH-----GTA 978
 DB 1188 LQPSGFLPEGSPFSLGOSQHDLSQSAFLSMLSTQSHNSLLSSSVLTPDSSPGKLD 1247
 QY 979 TEVAVETTPKQONL-----WF-----LQDS----- 1000
 DB 1248 PQSLEPEPEDEAQCPCPGQPFNFSAQIPCDAAPTPPPAVSEDQPTSLQLLASSKPM 1307
 QY 1001 ----- 1000
 DB 1308 NTFGAANPCSPVOLSSSTHLPQGTTPKRLSGDSEMSQSPGLGQPKRGRPPPKFFKQVEQ 1367
 QY 1001 -----QKELDELNCLHPQGIRESQLKXERLEKRYQDIHSI 1036
 DB 1368 HYLTLQTAQPIPEMCSGMMWIRDPETLDVLKALHPGIRKALHKLHLSK-HKDFLOEV 1426
 QY 1037 HLARKPNLGLKSCDGNQELLN--FLRSLIEVATRLQKGLGYVEETSSEFARVISLEKL 1094
 DB 1427 CL-----QPLTDPIFEPNEL-----PALEEGVMSWSPEKTYETDLAVLQW 1468
 QY 1095 KDFGECVIALQASVIKKFLOGFMAP----- 1119
 DB 1469 BELEQRVVLSDLOI-----RGWTCPTDSTREDLTYCEHLDPSPEDIPWRGRREGTVPO 1523
 QY 1120 -----KQKRLQSEDSAKTEEVDEEKKMVE-----EAK 1148
 DB 1524 RONNPLDLAVMLVLEQNVERRYLREPLWAAHVVEKALLSTPNGAPDGTSTISYE 1583
 QY 1149 VASALEKWTATREAQTFSRMHVLLGMLDACIKWMSAENARCKVCPPKGEDDKLLCDE 1208
 DB 1584 ITPRVVRWQTILERCRAAQVCLCMQGLERSIAWESKNVKVCLVCRKGDNDEFLLLCDG 1643
 QY 1209 CNKAHELCLRLYVDPDGEWQCPACQATARRNSRGRNYTEESASESDEDEDEE 1268
 DB 1644 CDRGCHYIYCHRPMBEAVPEGDWFCVCL----- 1671
 QY 1269 EEEEEEEDEYVAGLRPRKRTIRGKHSVIP---PAARSGRRPGKPKHSTRSQPKAP- 1324
 DB 1672 ---SQVVEEYTORPGFPKQK---RKSSFPLTTPEGDSRRR---MLSRSDSPAVPR 1721
 QY 1325 -PVDAAVDELVLTQKRSRRQSLELQKCEIILHKIVKYRFSWPFREPVTREAEYDV 1383
 DB 1722 YPEDGLSPK---RRHSMRSHHSDLTFCETILMEMESHDAAPFLEPNRLVSGYRRV 1778
 QY 1384 ITHRMDFTQVONKSCGYSYRQVEFLTMKQVFTNAEVVNCGRSHV 1429
 DB 1779 IKNPMDFTMRBLLRGYTSSEFAADALLVFDNCQTFNEDDSEV 1824
 RESULT 20
 Q8C8D1 PRELIMINARY; PRT; 718 AA.
 AC Q8C8D1;
 DT 01-MAR-2003 (TREMBLrel. 23, Created)
 DT 01-MAR-2004 (TREMBLrel. 23, Last sequence update)
 DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
 DE Mus musculus 10 days neonate cerebellum cDNA, RIKEN full-length
 DE enriched library, clone:1930060C03 product:hypothetical Bromodomain/
 DE PHD-finger containing protein, full insert sequence.
 GN Name=BC065123;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.


```
QY 1328 DAEVDELVLQTKRSSRRQSLQCEILHKIVKYRFSMPFRBPVTRDEADYDVITHP 1387
Db 2022 -----RDDSNDLAICSMILSELETHEDAMPFLPVNLKLVPGYKVKKP 2066

QY 1388 MDFQTQVONKSCGSYRSVQEFLLTDMKQVFTNAEYVN 1423
Db 2067 MDFSTIRDLKUTSGQYPNVEAFSLDVLVDFDNCFTEN 2102

RESULT 23
QY03S9 PRELIMINARY; PRT; 2060 AA.
AC QY03S9;
DT 01-MAR-2004 (TEMBLrel. 26, Created)
DT 01-MAR-2004 (TEMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TEMBLrel. 26, Last annotation update)
DE AgCP10911.
GN Name=agCG50231; ORFNames=ENGANGG00000009298;
OS Anopheles gambiae str. PEST.
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Anopheles.
OX NCBI_TaxID=180454;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PEST;
RA Anopheles Genome Sequencing Consortium;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
CC -! CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC EMBL; AAAB01008964; EAA12387.1; -.
DR HSSP; Q9UIG0; 1P62.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR001487; Bromodomain.
DR InterPro; IPR004022; DDT.
DR InterPro; IPR001739; Methyl-CpG_bind.
DR InterPro; IPR001965; Znf_PHD.
DR Pfam; PF00439; Bromodomain; 1.
DR Pfam; PF02791; DDT; 1.
DR Pfam; PF01429; MED; 1.
DR Pfam; PF00628; PHD; 2.
DR PRINTS; PR00503; BROMODOMAIN.
DR PROSITE; PS00014; BROMODOMAIN_2; 1.
DR PROSITE; PS00827; DDT; 1.
DR PROSITE; PS01359; ZF_PHD_1; UNKNOWN_2.
DR PROSITE; PS00016; ZF_PHD_2; 2.
SQ SEQUENCE 2060 AA; 228520 MW; CBEF2DDFF7C95D40 CRC64;

Query Match 6.1%; Score 487; DB 2; Length 2060;
Best Local Similarity 16.3%; Pred. No. 8.6e-12;
Matches 329; Conservative 258; Mismatches 538; Indels 894; Gaps 57;

QY 151 KSGDACSPPSSKENSQIAQHQKKTETVVEDEGRRESINRRARSPRK-LPTSLLKGE 209
Db 161 RSDSTHSASSSPSPASGANSLSQSLTTITAAAGGAAGNANDSSRSQFKEGPRNLGRGV 220

QY 210 RKAAPPKFLPHKYDVQKLNEDKTIISNVPADSLIRTERPKNKEIVRYFIHNLALRAGTGEN 269
Db 221 SK-----PK-----KNTVASLQKRAVGSKPL----- 243

QY 270 APWVVEDELVKYSLPSKFSFDLLDPYKYMTLNPSTKRKNTGSPDRKPSKSKTDNSLS 329
Db 244 -----TAQQLLTQDAEIEKLRQAMLEASRNQSDNNTSNTTDTESISGMSGESGE--- 297

QY 330 SPLNPKLWCHVLKSLSGSLPKVKNKSNKSPHEEEMKMGPNKLIHTNFHIPKGGP 389
Db 298 -----HINLKE-----LRV----- 306

QY 390 PAKKPKGKSHDKPLKAGRSKGIINGQKSTGNSKS-----PKGLKTPKTKMKQMTLLDWA 444
Db 307 -----PLEKGWRRETIVIRGLTRNGHIKGDVYYPPOSV-----NMMKGMNIQILY 351
```

```
QY 445 KGTQKMTAPRNSGGTPTRTSSKPH-----KHLPPAALHLIAYYKENKDREKRS 494
Db 352 LQDFK-----PKDLSRDNFSSAKAIVGTFLQAPL---PYATDGE-----P 390

QY 495 LSCVSKTARLLSSD-----RARLPBELSLVQRYELLEHKKWASMSSEQRKEYLK 548
Db 391 IKMTDVEVARRL--EDLKNMFTRHAGLVGEQRIEIAKQQQALRADAKLAKEEMNKKEK 448

QY 549 KKRE-----ELKKKLKEKAKER-----REKEMLER-----LEKQKRYED 582
Db 449 ERRRQHMALIKQLELRKFEKEKKHQVILDKLQREKLVNRKDDTWILAELRPOED 508

QY 583 QELTGKN-LPAPFLVDTPPEGLPNTLFGDVAMVVEFLSCYSGLL-----LPDAQYPI 635
Db 509 SEIVDQTVLPSPSRIPGLK-LGTGVDADLLMWFEFLHNFGETLGFDMESLPLNQLSLHAL 567

QY 636 SLMEALSADKGGFLYLNRLVILLQTLQTLQDEIAEDYGE-----LGMKLSBIPITLHS 691
Db 568 TSENAIDABE-----ELLSVMTHLLVCAIEDPGIPNPGRHTTLLGQTLRQADITHN 619

QY 692 VSELVRLCL--RRSDVQSESGSDTDNNKDSAAAPEDNEVDQEFLEKLET--SERPEL 746
Db 620 VSEILRIYLYAVATGEVKLQS--GINLERDRD--APSKHHLVNDDEDFKMCSTKXNSQFYEL 677

QY 747 EE-----KLQILTALCHRIIMTYSVQDHMETROQMGAELWKERL 785
Db 678 ENARYKLSLLKDKPPVALNPTTKTEILLAMLCNDLLMKNKAVCKQIDSSLEAQALAKERY 737

QY 786 AV----- 787
Db 738 LLDNKIRKYKMLSPVPMGEVGETPNGGVFNIPGKPLTAEMTNGGSEQPSHQOQTPQVH 797

QY 788 -----LKENDKK-RAEKOKRK- 803
Db 798 QNQSNSASSELDLLSKSMINDHSTCGDRAEDNSDLESEGTQLEDEDAHLTAEBAQRKY 857

QY 804 -----EMEAKNKENGKVE 816
Db 858 DKILETSFNKQOLENALNQLRVKCFQGDQRYWRYWNLGKCGGIYVEAMESIOPENYRYE 917

QY 817 NGLGKT-DRKKRIIVKPEQVDTE-----ADM-----ISAVKSRRLAIQAKK 858
Db 918 NALBEVQSRPDYVPKYAPAKPESAVKQDPAEDVAAECKKEGEVDVAVKPEIADAQGTK 977

QY 859 EREIOERE-----MKVKLEQAEEERIRKHAAAKAFQSGIAKA-----KL 901
Db 978 Q-ELQPTEDRKRCSSVSLKGLKKKKKQRTTSEGGESFRGNLGSAAATPDDDDVTVW 1036

QY 902 MRTP----- 906
Db 1037 QSETPTITIPDDDTESGDRMPTAPCEGVKTEHDNKGSTTLTDSMNCMDKPKLMENGVE 1096

QY 907 -----TGTDRNH----- 913
Db 1097 DFGELVEQQLQEVDDDEBIVTGARGNHGSANDCKPMIAAGDNGGVDRKPRTTTTSSVIS 1156

QY 914 -----NRYW----- 917
Db 1157 YGSOVEMIEVDEDSNCAVRAPITPGSVLYLRNTDTQKSEFSDQLLNRFPSIVDKELP 1216

QY 918 LPSDEVP-----GLFIEKG----- 931
Db 1217 LSTTECPLTINGSTPASLARQIPTNITCREICQIQGRNWDIGNNIQPSVPLEKGVETH 1276

QY 932 -----WVHDSIDYRFNHHCKD----- 947
Db 1277 FPNESILSMGLDDDBEINVIKAKSRPEPLQLSDMKPAFKRETI EYSQHPHQLRA 1336

QY 948 -----HTVSGDEDCPRSKKANLGKASMNT----- 978
Db 1337 EMMASETADPEYGNFSLPAYMTLTLNLTAIVQCDQFQPLQMTPEEKQLEDVQKHG-A 1395
```

```
QY 979 TEVAVETTTPKQGNLWFLCQSKDELINCLHPQIGRESOLKERLEKRYQDIITHSL 1038
Db 1396 POKTEPOVVPRFRYGMWKINDIELELNELIKALNPRGRVRLROSL---LESIAESVNL 1452
QY 1039 -----ARKPNLGLKSCDGNQELNIFLRS-----DLIE-----VATRLQKGG 1075
Db 1453 TTPHHVSHPRAPPNGYIEPEAWNAPSIARRVEVALDQIEAMEDKVASASQVKGW 1512
QY 1076 GIVVETSEFEARVISLEKLRQGECEVIALQASVIKKFLOGFNAPKQKRKL-----QSDS 1131
Db 1513 QMPQREGSGENVGVDVTIEMLRILGLEAAIERRYLKPPLGINTTTEAQMVAIAQOESH 1572
QY 1132 AKTEEVDE-----EKKMVEAKVASALEKWKTAIREAQTFSRMHVLLGLMDACIKWDMGAE 1187
Db 1573 SNQNVSNLSCNSASDENUPKGLLSWRDAVERSVTTAQSALYVLESVCVAMDKSIM 1632
QY 1188 NARCKVCPKGGDDKLLILDCENKAFHLFCLRPALYEPDGEWQCPAQO-PATARRN--- 1243
Db 1633 KANCQFCQSGESDGLLLCDGDCRGYHTYCFKPRMDKIPDGDWYCFECKNKATGDKCIV 1692
QY 1244 -----SRRNYTEE----- 1252
Db 1693 CGGLRPPPLGKMWYCELCPRAYHQDCYIPMLKYPRGKWKYCONCAKAPPKKKPKRKE 1752
QY 1253 -----SASEDSEDDSD----- 1264
Db 1753 RTNNSSOSLNSLNSNQSLNSHEDIATPLSATSAATVEQAQSDYVASGEAAQY 1812
QY 1265 -----EEEEEEEEEEEDVEAGLRPRKTRKTHSVIP----- 1300
Db 1813 SASEGYTHAQSTSECEPQOQSPAEVDQIEASSESPVPTAGGGEYFSPSSNANNAAGSS 1872
QY 1301 -----PAARSGRRPKKPHSTRSQPKAPPVDDA----- 1329
Db 1873 TSYLSDRTPGQDGESECGDSEYQPRSPVKDSLCLAGSSSSSSSSISDHRKER 1932
QY 1330 -----EVDLVLQTRKSSRQSLQKCEEILHKIVKRFSPFPREPVTROEAEDYDVI 1384
Db 1933 SKERDEAKERAQOEKATKRLKELAVCKTILEEMELHEDSPFLLPVNTKQFPTRYKI 1992
QY 1385 THPMDFQTVQNKCSGYSRQVQFELTDMKQVFTNAEVYN 1423
Db 1993 KSPMDLSITIKRLQDLVYKSRDFDIADVRQIFDNCIVEN 2031
ID O45075 PRELIMINARY; PRT; 1370 AA.
AC O45075;
DT 01-JUN-1998 (Tremblrel. 06, Created)
DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)
DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)
DE Hypothetical protein F53H1.4.
GN Name=F53H1.4; ORFNames=F53H1.4;
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxId=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=99069613; PubMed=9851916;
RG WormBase Consortium;
RT "Genome sequence of the nematode C. elegans: a platform for
RL investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Beck C., Kramer J., Keppeler D.;
RT "The sequence of C. elegans cosmid F53H1.";
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
```

```
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Waterscon R.;
RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RG WormBase Consortium;
RL Submitted (SEP-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF045641; AAC02578.2; -.
DR PIR; G88637; G88637.
DR HSSP; Q13263; 1FP0.
DR WormBase; WBGene00018778; F53H1.4.
DR WormPep; F53H1.4a; CE31539.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR001965; Znf_PHD.
DR Pfam; PF00628; PHD; 1.
DR SMART; SM00249; PHD; 1.
DR PROSITE; PS01359; ZF_PHD_1; UNKNOWN_1.
DR PROSITE; PS00016; ZF_PHD_2; 1.
KW Hypothetical protein.
SQ SEQUENCE 1370 AA; 158292 MW; 1E508414B13CC534 CRC64;

Query Match 5.8%; Score 463.5; DB 2; Length 1370;
Best Local Similarity 20.0%; Pred. No. Se-11;
Matches 319; Conservative 194; Mismatches 531; Indels 549; Gaps 59;

QY 71 EVAELLKEEPAPWVEKLVLEMHVHNTASLEKLVDTAWLEIMTKYAVGECDFEVGKERVL 130
Db 74 EIEDLIDKEFPAG-EKLMSGKEYTVVSSSEKRGGLTYTMEDGTGKIGHR---DLRRKGL 129

QY 131 KVKIVK---IHPLEKVDDEATEKSDGACDSP-----SSDKENSQIAQ----- 171
Db 130 SVEETIKIAIEDAEFVDEKQKVRBELAENPIREIKKYAPIFSANRKSSTPKTAQLSVA 189

QY 172 --DHQKQKTVKVEDGRRRESINDRARRSPKLPSTLKKGERKQWPKPLPHYKDYKLVQNE 229
Db 190 EADSDVQEVSMDAESGAAANKSAMKTPRGAP-----RASGGFVLSR---RLQEK 239

QY 230 DKIIISNPADSLIRTERPPNKEIVRYFIRHNALRAGTGENAPWVVEDELVKYSLPSKFS 289
Db 240 QKEXDEKEKLEKKQKEQ-----EKKQKKEEKAKK--LKEK-E 277

QY 290 DFLDDPYKYTLNIPSTKQNTGSPDRKPKSKSTDNSSLSPLNPKLWCHVHLKSLSGS 349
Db 278 EKLKEEKAARAKAEKKEKNGTMDK-----FLKXD-TGS 311

QY 350 PLKVKNSKN-----SKSPEEHLEEMMQWSPNKLHTNFHPKKGPPAKPKGKHSKPLK 403
Db 312 P-----SSKNAPLFSFKWGEKRIAVGVKKM-----EDAKRRDLLEY 349

QY 404 AKGRS---KGILNGQKSTGNSKSPKGLKTPKTKMKQMTLLDMAKGTQKMTRAPRNSGT 460
Db 350 NEACSWCEKNLSQGNRST-----FENPIKFSVQKLVDAK----- 385

QY 461 PRTSSKPHKLPAPPAALHUIAYKENKOREDKRSALSCVISKTARLLSSEDRARLPEELRS 520
Db 386 -----DRA----- 388

QY 521 LVQKEYELLEHKK--RWASMSSEQRKEYLKKREELKKLKEKAKERKEKMERLEKQK 578
Db 389 -----HMKGMKWAQKAE--FKAEMSEKRELYQKFPFKIAXWFNEDI----- 428

QY 579 RYEDQELTGNLPAP--RLVDTPEGLPNTLFGDVAMVVEFLSCYSGLLLPDAQYPIAVS 636
Db 429 ALDDLVTVCNLELLDAKRVNCDELLKCL-----EISQFFVSMRKILLNEN--ITAEQ 481

QY 637 LMEALSADKGGFLYNRVLVILLQTLLOTLODETAEDYBELGKMLSEIPLTLHVSSELV 696
Db 482 LRDDL---HGGFDGFKRSTYKMIANLLETALQEKHEKAHAHCNARLSSEPFINEITSELI 538

QY 697 RLCLRRS-----DVQSESESGSD-----TDDNKDSAAPE 724
```

Db 539 RAFFIGSTSTFKRDKRAARGHDDDDDEEDGDEEMDSREKVVIEEKEPENVNGAAE 598
Qy 725 DNEVQDEFLEKLTSEFFETSEKQLITLALCHILMTYSVQDHMETROQMSAEIWKKE- 783
Db 599 NGEIGDDEENSES-----ELQKRIILALPADSCHIELPAGAQ--LEVLCAW-----KEV 647
Qy 784 -----RLAVLKEENDKGAEKQK-----RKMEAKNKE-----NGKVENGLTKDRK 825
Db 648 VHGLPIIREWFLDANSEKLTTEHKEANRIRNEMEQHOQLQDPFPIPEITTESMTRTQTR 707
Qy 826 -----KRIVKPEPOVDTEADIMISAVKSRLLAIQAKKEREIQEREMKVKLERQAEER 879
Db 708 EASQSRREKLENDLDT-----LKI ELENREATAREVD-DLER----- 746
Qy 880 IRKHAAAEAKAFOEGIAKALVMRTPIGTDNRHNYWLF---SDEVP-----GLFIEKG 931
Db 747 -----IFRVVHIGNDRHLRKYYWYFAYSSDAIWWQDEGTTTSYEK 785
Qy 932 WYHDSIDYRPNHCKDHTVSGDEEDYCPRSKKNVGNASMTQHGTAATEVAVETTPKQG 991
Db 786 WVRDCSEKGF-----DVESSDVENRPEVEDLP-----TSSQS 819
Qy 992 QNLWFLCDSQKDELINCLHPGIBESQLEKRYQDIHSHLARKPNLGLKSCDG 1051
Db 820 SETWYKLDTEPAIRQLMTQKNGREKLLKYLNNMDDITSSILRKEKQ---KKDGG 876
Qy 1052 NQ-----LNLFLRSDL 1063
Db 877 EEEDEASEDEASAAENGEEKMETEQNETAAAAAETQEAENARRFTGRFGLKRTM 936
Qy 1064 IEVATLQGLGYVEETSEFEARVISLEKLDKFG-----ECVI---ALQASV 1108
Db 937 SELLNDWKQSGISKIVDSQVFARMLEANTLDEMGRVLTSLTIPVECVIEKFPQWAI 996
Qy 1109 IKKFLQGFMAKQKRKLQSED-----SAKT-----EEVDEEKWVEEAKV 1149
Db 997 AKKFLNFSTILLEKKNYEFSDFFRAKCRIFNYFLFSHKTSKFLEKLNQTKIGAKFGI 1056
Qy 1150 ASALEKWKTA-----IREAQTFSRMHVLLQMLDQACIKWDMASEN 1188
Db 1057 CAIFQKPKDVRLLPKKCECFSHLKMFRFCRRVQEAENASCLHWLLAYFARDIQORTLPE 1116
Qy 1189 ARCKVCPKGGEDKLLILDECNKAFHFLCLRP-----ALYE---VPDGEWQCPACQAPAR 1241
Db 1117 LSCQVCRKRTGTGERKLMKQCSTVFHYGCHRPITISRALFEEGFKBEG-WWCAKCTKEDRR 1175
Qy 1242 RNSRGRNYTEESASESD-----DESDEEEEEEEEEEEERYEYAGLRPRKTI 1292
Db 1176 R-----QLSEAKEDLRQKEGGDEEDHGGSGESDDEEDIVEEETRGRSAKR- 1224
Qy 1293 RKGHSVIPPAARSGRRPGKPHSTRSRQKAPP-----VDDAEVDLVLQTKRSRRQS 1346
Db 1225 -----KANAMRDVLEFEGVLROTAPPFPKQKTVVPEVLELNSIERANPRLY 1275
Qy 1347 LELQKCEELHKIVKYRFSWFPREPVRDEADYDVITHPMDFTQVQNKCSGYSRSVQ 1406
Db 1276 KTLQOI-----PGQSRSTRNAQHENRSLPDIEQDLV-----YTSAE 1312
Qy 1407 EFLTDMKQVFTNA-----EYVNCRGSHVLSMWK 1435
Db 1313 QLHEHLSQFFRHARGVIETHNPRKFSKFCPFK 1345

RESULT 25

Q23590 ID Q23590 PRELIMINARY; PRT: 1376 AA.
AC Q23590
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Flectin protein 1.
GN Name=flect-1; ORFNames=ZK783.4;
Query Match 5.6%; Score 449.5; DB 2; Length 1376;
Best Local Similarity 18.7%; Pred. No. 1.9e-10;
Matches 287; Conservative 247; Mismatches 55; Indels 447; Gaps 55;
Qy 126 KEKMLKXIVKIHPLKLEKVEATEKKSQACDS-----PSSDKENSSQ----- 168
Db 21 QQQLLQQLAKI---QKATASSPSKSTNGTSASTSAVSTSGTSSSQNEAAQLNLAKWQ 77
Qy 169 -----IAQDHQKKTIV-----KEDEGRRESINDRARSRKPLPTSLKKGERK 211

OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=99069613; PubMed=9851916;
RG WormBase Consortium;
RT "Genome sequence of the nematode C. elegans: a platform for
investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Favello A., Vaudin M.;
RT The sequence of C. elegans cosmid ZK783.";
RL Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Waterston R.;
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Wilson R.;
RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Wilson R.;
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RG WormBase Consortium;
RL Submitted (SEP-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; U13646; AAC24421.2; -.
DR PIR; T34516; T34516.
DR HSP; Q14839; 1MW3.
DR WormBase; WBGene0001470; flect-1.
DR WormPep; ZK783.4; CE34152.
DR GO; GO:0003677; R:DNA binding; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR001487; Bromodomain.
DR InterPro; IPR004022; DDT.
DR InterPro; IPR001739; Methyl-CpG_bind.
DR InterPro; IPR001965; ZnF_PHD.
DR Pfam; PF00439; Bromodomain; 1.
DR Pfam; PF02791; DDT; 1.
DR Pfam; PF01429; MBD; 1.
DR Pfam; PF0628; PHD; 2.
DR PRINTS; PR00503; BROMODOMAIN.
DR SMART; SM00297; BROMO; 1.
DR SMART; SM00571; DDT; 1.
DR SMART; SM00391; MBD; 1.
DR SMART; SM00249; PHD; 2.
DR PROSITE; PS50014; BROMODOMAIN_2; 1.
DR PROSITE; PS50827; DDT; 1.
DR PROSITE; PS50016; ZF_PHD_2; 1.
DR SEQUENCE 1376 AA; 154779 MW; 6D5E78C61D194203 CRC64;

Db 78 QIQQLAFCALMAAOKQOQKAAADKAKEKEKAKAAAAAASASTSSASAPG 137
Qy 212 WAPPKPLPHYDVKLQNEKIIISNVPADSLIRTERPPNKEIVRYFIRNALRAGTGENAP 271
Db 138 LSPLEAAWQAQAIQWQ-----ALQOMMTPOKSON-----167
Qy 272 WVVEDELVKYSLPSKSFDFLLDPYKMTLN--PSTKRTKNTGSPDRKPSKSKTD-----324
Db 168 ----BEAIIK-----MMDMAKKPAGVASTSSASTSTSTSSASTSSNNNA 213
Qy 325 NSGLSPLNPKLWCHVHLK-----KSLGSPKLVNNSKNSKSP 362
Db 214 NNAASNMNMVWQLVAAQMOQKQOQKQDQKQADQAKAKELAKQOQKEQDVKNKQ- 272
Qy 363 EHLBEMMK-MMSPNKLTNFIH PKGPPAKPKGKHSDPLKAKGRSGKILNQKSTGNS 421
Db 273 ----BEILFLMAQHOLN-----HOKKHEKKQADAAALAAKVLAAHRAALES 315
Qy 422 KSPKGLKTPKTKMQLTMDMAKGTOKMT-----RAPNSGGTPTSS 465
Db 316 DSPEGKKTNEAMLR-----LPQLGWRRTOTCVRSIASAGVGDVYFAP--CGKKLSTVS 369
Qy 466 KPHKLPALPAALHIAYYKENKDRKRSALSCVISKTARLLSSEDR---ARLPEELRSIV 522
Db 370 EVVRYLTGNSIHVIT--RDNFLNTKLVIGEIFVPKQTEADETQOEREBFAMFTEDDINKE 427
Qy 523 QKRYELLEH-KRWASMSBQKRYLVKKR-----BELKKLKE-----560
Db 428 LTRLNVLKFPVKIQASTSGNVHEDDIKMSKIEEPDEPLDPSSELNDEFTBELVHSIQMSNG 487
Qy 561 ----KAKERREMERLEKQRYEDQELTGKLPFAFLVDTPEGLPNTLFGDVMVVVF 616
Db 488 VDEKIRERADLLVINDV-----RHLPDFSRIGN--QCLSSQGFADALMVHBF 536
Qy 617 LSCYGLLPDQAPYITAVLSMEALSADKGGFLYLNRLVILLQTLQTLQDEIAEDYG 676
Db 537 VQNFHVLGIDLEIAPKLESKAGLDGDAN---HAEQTLQLTRQ--LLRLALEFPFGNGEK 592
Qy 677 ELGMLKSEIPLTLHVSVELVRLCLRRSDVOBESGSDTDNDKDSAAFENEVDQEFLEKL 736
Db 593 RFQGGGEMGLDRENFSEVWRLFL---IDKGRG-----EELSQPL 630
Qy 737 ETSEPELTSSEKLIJLTALCHRLMTYSV---QDHMETROOMSAELW-----781
Db 631 LTCNFUSIPEQKASLAFDELVCVRNVVTEIDKNLDEISRLKEKMRGKARLRS 690
Qy 782 -----KERVAVLEE-NDKRAEKQKKE-----MEAKNKENGKVE 816
Db 691 ARSKKNDEKVVVKEEQNHESDEPPTPTPKATVAPPTVSVSPVSAQQOQKPT 750
Qy 817 NLGKTDKRRIVKFPFQVDTBAEDMISAVKSRRLIAIOAKKEREIQEREMKVLERQAE 876
Db 751 PGLGQC-----EVLTEQESMSLQOMDSLIGDLHQBQAINQK-----788
Qy 877 EERIRKHAAKAEKAFQEGITAKAKLVMRRTPIGTDNRHNYWLF--SDEVPGLFIEKGWH 934
Db 789 -----IHDGLKIRSPFGDTRFRNYWMLATDKV-----IIESLAT 826
Qy 935 DSIDYRFNHCKDHTVSGDEDYCPRSKKNALGNKASMTQHGTAATEVAVETT-----986
Db 827 TSVN---NPACNANAYSKDP--FTLEQRVPACETIDLDVIACVEDLDVDDVLLRAKAD 881
Qy 987 --TPKOGQNL-----WFLCDSQKELDELLNCLHPQIGRESOLKRLKRYQDIHSHI 1037
Db 882 KKTREKYRRIENHMKRGWMTQNRDCVESLRSCMLSRGIRERALLRLTKPW-----933
Qy 1038 LARKPNLGLKSCDGNQELLNFRSLDIEVATRLQKGLGVVEETSEFEARVLSLEKLKDF 1097
Db 934 -----FNLKFGTITIEPVEKESDLDLVRQGWTRLNT-AIDKL-----972
Qy 1098 GECVIALQASVIAKKFLQGF-----MAPKQKRLKQSDSA-----KTEEVDE 1139
Db 973 -QC--HLKMSDVSXPLPSITPTPTQKPIVVVPTMALAQIVKDDMAWKVDEEVGQDELDE 1029

Qy 1140 ---EKKVVEEAKV-----ASALEKWKTAI-REAQTFSRMHVLLGMLDACIKWDMSAENARC 1191
Db 1030 TTIROKIIETADMDODTQOLFEDMKSVYSTEAQTTSQLMVALQTUEGIMMWSREALC 1089
Qy 1192 KVCPPKGGDDKLIILDECNKAFHLFCLRPALYVEPDGEGWQCPACQAPATARRN-----1243
Db 1090 QIC-KSMGDDEMLVCDGCGESGCHMECFRPMTKVPEDGDFCQRCREKSGRPMCFCSRE 1148
Qy 1244 -----SRGRNYTESASESDESEDEE-----EEEEEE 1274
Db 1149 TGNLHQCQKCAHYHVQEGSQDQPKCAINPETPFCGHCOEMKQMRVFKRLLILASESEREL 1208
Qy 1275 EEDYEVLARLRPRKTIRKSHSVIPPAARSG---RRPGKPHSTRSQPKAPPVDDARV 1331
Db 1209 EDDNHAENG-----ENTKNHNGMNGAIAIGVHNQONGVGNLKRK-----LEV 1253
Qy 1332 DELVLTQKRSRRQSLQKCEEILHKIVKYRFSFPFREPVTTRDEADYDYVITHPMDFQ 1391
Db 1254 PSI-----GGLPKMNKELCQLMDLVLVQANALPFLEVPNPKLVPGVKMIISKPMDLK 1307
Qy 1392 TVQNKCSCGSRYSVOEFLTDMKQVPTNAEVNCRGSHV 1429
Db 1308 TIRQKNEKLIYETPDFAEDIELMPANCROFNIDHSEI 1345
RESULT 26
BAIA XENLA
ID BALA XENLA STANDARD; PRT; 627 AA.
AC Q8UVR5;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE ATP-utilizing chromatin assembly and remodely factor 1 (xACFI)
DE (Fragment).
GN Name=ACFI;
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21668941; PubMed=11809820; DOI=10.1091/mbc.01-09-0441;
RA Macallum D.E., Losada A., Kobayashi R., Hirano T.; identification as
RT "ISWI remodeling complexes in Xenopus egg extracts: identification as
RL Mol. Biol. Cell 13:25-39(2002).
CC -!- FUNCTION: May play a role in transcriptional regulation.
CC -!- SUBUNIT: Together with p18 and p20 proteins, it forms the xenopus
CC version of CHRAC.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- PTM: Phosphorylated in mitosis.
CC -!- SIMILARITY: Belongs to the WAL family.
CC -!- SIMILARITY: Contains 1 bromodomain.
CC -!- SIMILARITY: Contains 1 PHD-type zinc finger.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF412332; AAL60160.1; -.
DR HSSP; Q9UIG0; 1F62.
DR InterPro; IPR001487; Bromodomain.
DR InterPro; IPR011011; FYVE_PHD_Znf.
DR InterPro; IPR001965; Znf PHD.
DR Pfam; PF00439; Bromodomain; 1.
DR Pfam; PF00628; PHD; 1.
DR PRINTS; PR00503; BROMODOMAIN.


```

Db 857 LAALKER---LEKAEKRIAEKRIAEENKRIAEKRIAEEL-----EKKRLQK----- 904
Qy 836 DTEADMI SAVK-SRLLAIQAKREIQEREMKVKLERQAE-BRIRKHKAABAKAQE 893
Db 905 --EQDRLAAELERKLEKAEKRIAEKLEKK-LEKAEAEVKRIADEAAAAAKLEKE 961
Qy 894 GIATAKLVMRPTIGTDNRHNRWLFSDVPGFTIEKGWVHDSIDYRFNHCKDHTVSGD 953
Db 962 -----RLEKAEAEKRIADE 975
Qy 954 EDYCPRSKANLGNAMNTOHGTATEVATTTTPKQGNLWFLCDSQKEL-----DELL 1008
Db 976 AAAAAKLEKLEKAEKRIADEAAAAEAKLEKE-----RLEKAEKRIADEAAAAA 1029
Qy 1009 NCLHPGCI-RESQLERLEKRYQDIHSHLARKNLGLKCDGQOELLNFRSLDIEVA 1067
Db 1030 AALLOQKI EKEKEERDRIAKENKELKEKEDKERQORERQEKQERARALKIEKEK 1089
Qy 1068 TRLOKGLGVRETSFEARVISLEKLDKPGECVIALQASVIKKFLQGFMAPKQRRKLO 1127
Db 1090 ERLNQKLDQEKERERQERKEQOEREENE-----KQLEKEREKEREK 1138
Qy 1128 SEDSAKTEVEDEKMWEEAKVASALEKWK-----TAIREAQTFSRMHVLLGML 1176
Db 1139 QRNEQLEKERQERFKDQEKQKLEKEQOQKQKVIPTTTTIRERS----- 1186
Qy 1177 DACIKWMSAENARCKVCPKGEDDKLILCDECNKAFHLCURPALYEVDPDQEQPACQ 1236
Db 1187 -----NSDSALLNLPDSASSSH-----SELVHLTLGRSAKGRK 1222
Qy 1237 PATARNSGRNVTESASEDSE-----DDSDSEEEEREEREEDYEAGL--- 1284
Db 1223 KPTRRELTQGNRNKNINLNLKELPSTVIIIPDDDDQSQKQOQSEVEEKPVAVKMPIG 1282
Qy 1285 ----RLRPR---KTIRGKHSVIPPAAARSRRPGKKPH-----STR 1317
Db 1283 FNPMMMPKFDPSAVKLKSSAKPPSTNTTTTTPAQPVQNLKPVALKPTVTOTTTT 1342
Qy 1318 RSQKAPPVDDAEV 1331
Db 1343 TTPPTTPPSSSVQL 1356

RESULT 28
Q8ISF5 PRELIMINARY; PRT; 10578 AA.
AC Q8ISF5;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE lmda_1 protein.
GN Names=iso1;
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22269627; PubMed=12381307; DOI=10.1016/S0022-2836(02)00970-1;
RA Flaherty D., Gernert K., Shmeleva N., Tang X., Mercer K.,
RA Borodovsky M., Benian G.;
RT "Titins in Caenorhabditis elegans with Unusual Features: Coiled-coil
RT Elastic Regions."
RL J. Mol. Biol. 323:533-549(2002).
RL EMBL; AV130758; AAN61519.1; -.
DR PIR; E89066; E89066.
DR PIR; T33247; T33247.
DR HSP; Q10466; 1BPV.
DR InterPro; IPR008938; ARM.
DR InterPro; IPR003961; FN.III.
DR InterPro; IPR008957; FN.III-like.
DR InterPro; IPR007110; Ig-like.

```

```

DR InterPro; IPR003598; Ig c2.
DR InterPro; IPR008941; TPR-like.
DR InterPro; IPR011046; WD40-like.
DR Pfam; PF00041; fn3; 2.
DR Pfam; PF00047; Ig; 3.
DR SMART; SM00060; FN3; 2.
DR SMART; SM00408; IGC2; 3.
DR PROSITE; PS0853; FN3; 2.
DR PROSITE; PS0835; IG_LIKE; 11.
SQ SEQUENCE 10578 AA; 1174295 MW; 71BF50C502FB4F16 CRC64;

Query Match      4.5%; Score 361.5; DB 2; Length 10578;
Best Local Similarity 20.2%; Pred. No. 8.2e-06;
Matches 294; Conservative 209; Mismatches 543; Indels 413; Gaps 60;

Qy 25 IPHQEAPRTREEVEARLERYSERIWTCKTSGS---SOLTKEAWEREQVEAELKEPFP 81
Db 8976 IGHVQEKTFATIELVFN-EFYAFRIVAVNGFEGAPSEIIEVNTLDYDQE-----ESPD 9028
Qy 82 -AWYEKLVLEWVHNTASLEKLVDTAWLEIMTKYAVGECDFEVCGEKMLVKVIKIHPL 140
Db 9029 FAGEELKLDVVQVNN-----EVVEITI-ESEVTIEEHRKLKK----- 9068
Qy 141 EKVDEBEATEKKSOGACDPSFSDKENSQIAQDHQKQKTVVKEDEGRRESINDRARRSPK 200
Db 9069 -----SKSKKTTDEPDLSEIALEVSSDITSSLEITTE-----STIPDTAPESQET 9115
Qy 201 LPTSLKGE---RKWAPPKFLPHKYVDVKLQNEDKTIISNVPADSLIRTPPNKEIVRYFI 257
Db 9116 LNVEIAVTETTVQKITNPDSAKDV---NEDTAVS-----SIVKK----- 9154
Qy 258 RHNALRAGTGENAPWVVEDELVKVYSLPSKPSDFLLDPYKYMTLNPSTKRNKTGSPDRKP 317
Db 9155 -----DDKDVNKKSLPES-----GLITTKLEIQGPEKKI 9183
Qy 318 SKSKSTDNSSLSPLNPKLWCHVHLKLSLGSPLKVNKSNKSPPEEHEEMKMWSPNK 377
Db 9184 MKKTEKADSSISSETSETLTQDLTQTKQSEPEPAK---RTTETSVDQEVKRTTETTSK 9240
Qy 378 LHTNPHIPKPPPAKPGKHSKDKPLKAGKSGILNGKSTGNSKSPKGLTKPTKMQ 437
Db 9241 QTTEEH-----PQGGKSDSSISSTDSASEVKVQVQSESEAOKVTEPATAKLESK- 9291
Qy 438 MTLDDMAKGTOKMTR-APRNSGTPRTSSKPKHKLPPAALHLIAYYKKNKOREDKRSALS 496
Db 9292 -----SKMTEDTTKESDNKETVDEKPKKV-----LKKKTESKDS----- 9326
Qy 497 CVISKTARLLSSEDRARLPPEELRSIVQKRYELLEHKRWASMESEQR---KEYLKKKREEL 554
Db 9327 -TISSETSAVESAGPSESETQNV-----AVDKKQKQKQKQKQKLEAEIAGKKSTEQ 9380
Qy 555 KKKLKEKAKEREKEMLERLEKQRYEDQELTGKVLPAFRL-VDTPEGLPNTLFGDVAMV 613
Db 9381 KSKLEAEAKLRAABE-DAAKQKQ--EKTEAASKAAAEKLEKLEKQQAQINKAAAEADAVKK 9437
Qy 614 VEFLLSCYSGLLLPDAQYPTITAVSL-MEALSADKGGFLYLNRLVILLQTLQLQDETA 672
Db 9438 QNELDEQNKL---EATKKLAELKLEEQSRAK-----SKQA 9472
Qy 673 EDYGLGMKLSLPIPLTHSVSELVRLCLRRSDVQSEBSGSDTDD--NKDSAAFEDNEVOD 730
Db 9473 EQAKLDAQ-----TKAAAEKQTKLEKDEKSNKDSGS---NETVE 9510
Qy 731 E-----FLEKLETSEFFELTSSEKQLIITLALCHRLMITYSVQDHMETR 773
Db 9511 EKPKKVLKKKTEKSDSSISQKSDTSK---TVAESAGSSESETQKADATSKQKQETDKK 9566
Qy 774 QQMSAELWKERLAVLKEENDK-----KRAEKQKRWEMAKNKGKNGVGLGKTDK 825
Db 9567 QKLEAEITAKKSA-----DEKSKLETESKLIKAAEDAAKQKQKEDKLEADVASKAAAE 9623
Qy 826 KRIVKFEQVDTAEADMISAVKSRLLAIQAKEREIOEREM---KVKLERQ-----AE 876

```

```
Db 9624 KLELEKQAIKKAEE--ADAVKKQKLEAEKQKLESEAAATKAAAEKLEEQQAQINKAAE 9681
Qy 877 EERIRKH-----KAAAE--KAFQEGIAKAK-LVMRRTPIGTDNRHNYLWFS 920
Db 9682 ADAVKQKELDEKNKLEANKSAEAKKLEESAASAKQTVBEQAQKDAQTKREKT---A 9737
Qy 921 DEVPGLFIEKGWVHDSIDYRNFHHCXHTVSGDEDYCPRSK--KANLGNASNNWTHGTA 978
Db 9738 EKQTGL--EKD-----DKSTKDSKSETVDEKPKKKVLLKKTTEKSDSISQKSVT 9785
Qy 979 TEVAVETTPKOG--ONLWFLCDSQKELDELLNCLHPQGIRESQLEKLEKRYQDIIHSI 1036
Db 9786 SKTVESGSPSESETQKVDAARKQKETDE-----KQLEAEI----- 9823
Qy 1037 HLAKEPNLGLKSCDGNQEL---LNFRLSDLIIEVATRLQKGLGVBEETSEFFARVISLEK 1093
Db 9824 -----TAKGSADESKLEAESKLAKEAEVEAAKK-QKEQBQLKLDTEAAKKAEEK 9875
Qy 1094 LKDFGECVIALQASVTKKFLQGFMAPKQK-----RRKQSEDSANTEVEDEEKQWVEAKV 1149
Db 9876 LE-----LEKQAQ-IKKAEEADAVKKEKELAEKQKLESE--AATKKAEEKLEKLEQKK 9926
Qy 1150 ASA-----LEKWTATREACTFPMHVLGMLDACIKWMSAENARCKVCPKGEDD--K 1202
Db 9927 KQAEYASIEKQEQEKLEAQEQKLEVDAAKSAEKQKLESETSKKTEAPKESVDEKPKK 9986
Qy 1203 LILDCBCKNAFLCLRPALYEPDGEWQCPACQATARNRSGRNYTBESASESE--- 1259
Db 9987 KVLKKTKEK-----DSSISQKSDTAKTVASAGQSDSETQK 10023
Qy 1260 -----DBSDSE-----EEEEEEEEEEEDYE 1280
Db 10024 VSEADKAHRQKESDEKQKLESEIAAKSAEQAESKLETEAKTKKVIDESAQKQEQEDKK 10083
Qy 1281 VAGLRLRPKTKTIRGHSIVPPAARSGRRPGKPHSTRSQPKAPPVDDAEVDELVLQOT-- 1338
Db 10084 KGDSANQKQDKQEKQKLESEAT-----SKRPTSEKQDKETPQBAKSENETVMTTEP 10137
Qy 1339 -----KRSSRRQSLE 1348
Db 10138 QQLEVKSEPKSKDKTETVE 10156

RESULT 29
Q8ISF6
ID Q8ISF6 PRELIMINARY; PRT; 18519 AA.
AC Q8ISF6;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE 2MDA.2 protein.
GN Name=iso1;
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22269627; PubMed=12381307; DOI=10.1016/S0022-2836(02)00970-1;
RA Flaherty D., Gernert K., Shmeleva N., Tang X., Mercer K.,
RA Borodovsky M., Benian G.;
RT "Titins in Caenorhabditis elegans with Unusual Features: Coiled-coil
RT Domains, Novel Regulation of Kinase Activity and Two New Possible
RT Elastic Regions."
RL J. Mol. Biol. 323:533-549(2002).
CC -!- SIMILARITY: Belongs to the Ser/Thr protein kinase family.
DR EMBL; AY130758; AAN61518.1; -.
DR PIR; B89066; E89066.
DR PIR; T33247; T33247.
DR HSP; Q10466; LBPV.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0004674; F:protein serine/threonine kinase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
```

```
DR GO; GO:0006468; P:protein amino acid phosphorylation; IEA.
DR InterPro; IPR008938; ARM.
DR InterPro; IPR003961; FN III.
DR InterPro; IPR008957; FN-III-like.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003598; Ig_c2.
DR InterPro; IPR011009; Kinase like.
DR InterPro; IPR006025; Pept M_Zn_BS.
DR InterPro; IPR000719; prot_kinase.
DR InterPro; IPR002290; Ser_thr_kinase.
DR InterPro; IPR008271; Ser_thr_kin_AS.
DR Pfam; PF00041; fn3; 11.
DR Pfam; PF00047; ig; 6.
DR Pfam; PF00069; pkinase; 1.
DR SMART; SM00060; FN3; 11.
DR SMART; SM00408; Igc2; 16.
DR SMART; SM00220; S_TKC; 1.
DR PROSITE; PS00583; FN3; 11.
DR PROSITE; PS00835; IG LIKE; 38.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS00111; PROTEIN KINASE_DOM; 1.
DR PROSITE; PS00108; PROTEIN_KINASE-ST; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
KW ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.
SQ SEQUENCE 18519 AA; 2051739 MW; 6A8441C5D0BA7729 CRC64;

Query Match 4.5%; Score 361.5; DB 2; Length 18519;
Best Local Similarity 20.2%; Pred. No. 1.6e-05;
Matches 294; Conservative 209; Mismatches 543; Indels 413; Gaps 60;

Qy 25 IPHTQEAFTREYERARLERYSERIWTCKSTGS---SOLTHKEAMWEEQEAELLKEEFP 81
Db 8996 IGHVQEKTFATIELVFN-EFYAFRIVAVNGFGEGAPSEIIEVNTLDYDQE-----ESPD 9048
Qy 82 -AWYEKLVLEWYHNTASLEKLVDTAWLEIMTKYAVGEECDDEVGKEMLVKIVKIHEL 140
Db 9049 FAGEBELKLDVQVANN-----EVVTEITI-ESEVTEIEHRKLKK----- 9088
Qy 141 EKVDEEATEKSGDAGCDSPSDKENSQIAQDHQKKEKTVVKSDEGRRESINDRARSPRK 200
Db 9089 -----SKSKKTTTDEPDLSEIALEVVSSDITSSLEITTE-----STIPDAPSESQET 9135
Qy 201 LPTSLKKGES---RKWAPPKFLPHKYDVKLQNEBKIIISNVPADSLIRTERPPNKEIVRYPI 257
Db 9136 LNVEIAVTETTVQKITNPSDESAAKDV---NEDTAVS-----SIVKK----- 9174
Qy 258 RHNAIRAGTGENAPWVVEDELVKYISLPKFSDFLLDPYKYNWTLNPSTKTKGTGSDPRKP 317
Db 9175 -----DDKVNKKSLPES-----GLTTKKEIQGKPEKKI 9203
Qy 318 SKKSKTDNSSLSSPLNPKLWCHVHLKSLSGSPLKVKNSKNSKSPHEHLEEMKQMSPNK 377
Db 9204 MKKTEKADSSISSETSETLTQDLOTQTKOSEPPAK---RTTSTVSQDEVKRTTETTSKSK 9260
Qy 378 LHTNFHPKPGPAKKPGKSHSDPKLAKGRSGILNGQKSTGNSPKKGLTKPTKMKQ 437
Db 9261 QTTEEH-----PQPGKSDSSISSTSDASEVQVQOSEAQVTEKPEAKLESK- 9311
Qy 438 MTLDMAGTKQWTR-APRNSGGTPTRTSSKPHKHLPPAALHLIAIYYKENKREDKRSALS 496
Db 9312 -----SKWTEDTTKESDNKETVDEKPKKKV-----LKKTKESDS----- 9346
Qy 497 CVISKTRALLSSDRARLPEELRSIVQKRYELLEHKRWASWSEQR---KEYLKKKREEL 554
Db 9347 -TISTSETSVASAGPSESETQNVN-----AVDKKKQKETDEKQKLEAEIAGKASTQ 9400
Qy 555 KKKLKEKAKERREKEMLERLEKQRYEDQELTGKQLPAFRL-VDTPEGLNPTLFGDVANV 613
Db 9401 KSKLEAEAKLKRAAES-DAAKKQK--EKTERAASKKAAAEKLEKQAQINKAAEADAVKK 9457
Qy 614 VEFLLSCYGLLLPDAQYPIITAVSL-MEALSADKGGFLYLNRVLLQTLLOTLLQDETA 672
Db 9458 QNELDEQNKL---EATKKLAEEKLEEQSAAK-----SKQAA 9492
```

```
QY 673 EDYGLGKMLSLPIPLTHSVSELVRLCLRRSDVQSESGSDTDD--NKDAAFEDENEVD 730
Db 9493 ESQAKLDAQ-----TKAAAEKQTKGLEKSNKDSGS---NETVE 9530
QY 731 E-----FLEKLETSEPELTSSEKLIQILTALCHRIILMTSVQDHMETR 773
Db 9531 EKPKKVLLKKTEKSDSSISQKSDISK-----TVAESAGSSESETKVADATSKQKETDK 9586
QY 774 QMSAELWKEKRLAVLKEENDK-----KRAEKQKRMKEMAKNKENGKVENGLGKTRDK 825
Db 9587 OKLEAEITAKGSA--DEKSKLETESKLIIKAAEDAAKKQKEDKLEADVASKAAAE 9643
QY 826 KRIVFEQVDTEADMISAVKSRLLAIQAKEREIQEREM---KVKLEQ-----AE 876
Db 9644 KLELEKQAIKKAAB--ADAVKQKELAEKQKLESEAAKKAABKLEEQAIKKAAB 9701
QY 877 EBRIRKH-----KAAAE--KAFQGGIAKAK-LVMRRTPIGTDRNHNRYLFS 920
Db 9702 ADAVKQKELDEKNKLEANKSAAEKKLEESAASKQTVVEQAKLDAQTKET----A 9757
QY 921 DEVPLGFIEKGWVHDSIDYRHNHCKDHTVSGDEDYCPRSK--KANLGNKNAASMTQHGT 978
Db 9758 EKQTKGL--EKD-----DKSTKDSKSEKTVDEKPKKVLKXKTEKSDSSISQKSVT 9805
QY 979 TEVAVETTPKQG--QNLWFLCDSQKELDELNCLHPQGIRESQIKERLEKRYQDIHSI 1036
Db 9806 SKTVESGSPSESTQKVADAAARKQKTD-----KQKLEAEI----- 9843
QY 1037 HLAKPNLGLKCDGNOEL--LNFRLSDLIIVATRLQKGGIYVEETSEFEARVISLEK 1093
Db 9844 -----TAKSADSEKSKLEAESKULKAAAEVBAKK-QKEKDEQKLDTEAAKKAABK 9895
QY 1094 LKDFGECVIALQASVIKFLQFMAPKQK-----RRKLOSEDSAKTEEYDEKRMVVEAKV 1149
Db 9896 LE-----LEKQAO-IKKAABADAVKKEKELAEKQKLESE--AATKKAABKLEEQK 9946
QY 1150 ASA-----LEKWTATREAGTFSRMVLLGMLDACIKWMSAENARCKVCPKGGED---K 1202
Db 9947 KDAETASIEKQEQEKLAQEQSKLEVDAKKSAEKQKLESETSKKTEEAPKESVDKPKK 10006
QY 1203 LILDCENKAFHLFCLRPALYVPPGEMQCPACQATARRNSRGNRYTEASASE--- 1259
Db 10007 KVLKKKTEK-----DSSISQKSDAKTVAESAGQSSETKQ 10043
QY 1260 -----DBSDE-----EEEEEEEEEEEDYE 1280
Db 10044 VSEADKAHKQKESDEKQKLESIAAKSAEQSKLETEAKTKVIEDESAKKQEQEDKK 10103
QY 1281 VAGLRPRKTTIRGHXSVIPPAARSGRRPGKPHSTRSOPKAPVDDAEVDLVLQT--- 1338
Db 10104 KGDDSAKQKQKQKQKQKLESEAT-----SKKPTSEKQKDEKTPQEKAKSENETYMTTEP 10157
QY 1339 -----KRSRRQSLE 1348
Db 10158 QLEVKSEPKSKDKTETVE 10176

RESULT 30
Q81SF7
ID Q81SF7 PRELIMINARY; PRT; 18534 AA.
AC Q81SF7;
DT 01-NAR-2003 (TrEMBLrel. 23, Created)
DT 01-NAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-NAR-2004 (TrEMBLrel. 26, Last annotation update)
DE 2MDA.1 protein.
GN Name=igof.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
```

QY 438 MTLDDMAKGTQKQWTR-APRNSGGTPTRTSSKPHKHLPPAALHLIAYYKKNKORDEKKSALS 496
Db 9312 -----SKMTEDTTKESDNKETVDEKPKKKV-----LKKKTEKSDS----- 9346
QY 497 CVISKTARLLSSDRARLPPEELASVQKRYVELLEHKKRWASMSSEQR---KEYLKKKEEL 554
Db 9347 -TISETSVAVESAGPSESSETONVA-----AVDEKKKQKETDEKQKLABIAGKSKSTEQ 9400
QY 555 KKKLKEKAKERREKEMLEKQRYEDQBELTKNLPAFRL-VDTPEGLPNTLFLGVDAMV 613
Db 9401 KSKLEAEAKLKRAAE-DAAKKQK--EKTEAASKAAAEKLEKQKQINKAAEADAVKK 9457
QY 614 VEFLSYSGLLLPDAPQYPITAVSL-MEALSADKGGFLYLNRLVLIQTLQLLODETA 672
Db 9458 QNELDEQNKL-----EATKLAEEKLEEQSAK-----SKQAA 9492
QY 673 EDYGELGMLKEIPLTLHLSVSELVRLCLRRSDVQEESEGSDTD--NKDSAAFEDNEVD 730
Db 9493 EEOAKLDAQ-----TYKAAEKQTLGKDEKSNKDSGS---NETVE 9530
QY 731 E-----FLEKLETSEFFELTSEELQILITLALCHRLMTYSVQDHMETR 773
Db 9531 EKPKKVLKKTEKSDSSISQKSDTSK---TVAESAGSSESETKVADATSKQKETDKK 9586
QY 774 QMSAELWKLRLAVLKEENDK-----KRAEKQKRKEMEAANKENGKVGNGLTKDRK 825
Db 9587 QKLEABITAKSA--DEKSKLETESKLKAAEDAAKQKQEKEDKLLEADVASKKAAAE 9643
QY 826 KRIVKEPEOVDTAEADMISAVKSRLLAQAQKEREIQEREM---KVKLERQ-----AE 876
Db 9644 KLELEKQAQIKKAAE--ADAVKKQKLEAEKQKLESAATKAAAEKLEEQQAQINKAAE 9701
QY 877 EERIRKH-----KAAAE--KAFOEGIAKAK-LVMRRTPIGTDRNHNRYLWFS 920
Db 9702 ADAVKKQKELDEKNKLEANKSAAEKLEESAAKSQTVEQAKLDAQTEKT-----A 9757
QY 921 DEVPGFLFIKGVVHDSIDYRFNHCHDHTVSGEDYCPRSK--KANLGKNASNNTOHGA 978
Db 9758 EKQTGL--EKD-----DKSTKDSSEKSTVDEKPKKKVLKKKTEKSDSSISQKSVT 9805
QY 979 TEVAVETTTPKQG--QNLWFLCDSQKELDELLNCLHPQIGRESQKLEREKRYQDIHSHI 1036
Db 9806 SKTVBESGGPSESSETKVADAAARKQKETD-----KQKLEAEI----- 9843
QY 1037 HLAKEPNLGLKSCDGNQEL---LNFRLSDLIEVATRLQKGLGYVEETSEFARVLSLEK 1093
Db 9844 -----TAKKSADKSKLEAESKLKAAAEVEAAK-QKEDEQLKLDTEAAKSKAAAEK 9895
QY 1094 LKDFGECVIALQASVIKKFLOGMAPKQK-----RRKLQSEDSAKTEBEVDEKQKVEEAKV 1149
Db 9896 LE-----LEKQAAQ-IKAAAEADAVKKEKLEAEKQKLESE--AATKAAAEKLEKLEQKK 9946
QY 1150 ASA-----LEKWKTAIREAQTFSRMHVLLGMLDACIKWDMSAENARCKVCPKKGEDD---K 1202
Db 9947 KQAEIASIEKQEKQKLEAQEQSKLEVDAAKSAEKQKLESETSKTKTEAPKESVDEKPKK 10006
QY 1203 LILCDECNKAFHLFLCLRPALYEPDGEWQCPACOPATARRNSGRNRYTESASEDSE--- 1259
Db 10007 KVLKKTESK-----DSSISQKSDTAKTVASAGQSDETK 10043
QY 1260 -----DQESDE-----EEEEEEEEEDYE 1280
Db 10044 VSEADKAHQKQESDEKQKLESEIAAKSAEQAQSKLEATEAKTKVI EDESAKQKEQEDKK 10103
QY 1281 VAGLRPRKTIKSHVIPPAAARSGRRPKKPHSTRSQPKAPFVDDAEVDELVLQT-- 1338
Db 10104 KGDSSAKQKQKQKQKLESEAT-----SKPTSEKQDEKTPQEKAKSENETVMTTEP 10157
QY 1339 -----KRSRRQSLE 1348
Db 10158 QQLEVKSEPKSKDKTETVE 10176

RESULT 31

Q80U42 ID Q80U42 PRELIMINARY; PRT; 886 AA.
AC Q80U42; AC
DT 01-JUN-2003 (TEMBLrel. 24, Created)
DT 01-JUN-2003 (TEMBLrel. 24, Last sequence update)
DT 01-MAR-2004 (TEMBLrel. 26, Last annotation update)
DE MKIAA0314 protein (Fragment).
GN Name=Baz2a; Synonyms=mkIAA0314;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=22579291; PubMed=12693553;
RA Okazaki N., Kikuno R., Ohara R., Inamoto S., Aizawa H., Yuasa S.,
RA Nakajima D., Nagase T., Ohara O., Koga H.;
RT "Prediction of the coding sequences of mouse homologues of KIAA gene:
RT II. The complete nucleotide sequences of 400 mouse KIAA-homologous
RT cDNAs identified by screening of terminal sequences of cDNA clones
RT randomly sampled from size-fractionated libraries.";
RL DNA Res. 10:35-48 (2003).
DR EMBL; AK122243; BAC65525.1; -.
DR HSSP; Q9UIG0; 1F62.
DR MGD; MGI:2151152; Baz2a.
DR GO; GO:006338; P:chromatin remodeling; TAS.
DR InterPro; IPR000637; A+T_hook.
DR InterPro; IPR001487; Bromodomain.
DR InterPro; IPR010111; FYVE_PHD_ZnF.
DR InterPro; IPR001965; ZnF_PHD.
DR Pfam; PF02178; AT_hook; 2.
DR Pfam; PF00439; Bromodomain; 1.
DR Pfam; PF00628; PHD; 1.
DR PRINTS; PR00503; BROMODOMAIN.
DR SMART; SM00384; AT_hook; 2.
DR SMART; SM00297; BROMO; 1.
DR SMART; SM00249; PHD; 1.
DR PROSITE; PS50014; BROMODOMAIN_2; 1.
DR PROSITE; PS50016; ZF_PHD_2; 1.
FT NON_TER 1
SQ SEQUENCE 886 AA; 10060 MW; 43E1172F5F081E06 CRC64;

Query Match 4.4%; Score 353; DB 2; Length 886;
Best Local Similarity 21.8%; Pred.No. 1,1e-06;
Matches 114; Conservative 78; Mismatches 188; Indels 144; Gaps 14;
QY 979 TEVAVETTTPKQGNLWFLCDSQKELDELLNCLHPQIGRESQKLEREKRYQDIHSHIHL 1038
Db 408 TQTAQPIPPCMSCGMMWIRDPE-TLDVLLKALHPRGIREKALHKHLSK-HKDPLQEVCL 465
QY 1039 ARKPNLGLKSCDGNQELN--FLRSDLIEVATRLQKGLGYVEETSEFARVLSLEKLD 1096
Db 466 -----QPLTDPIFEPNEL-----PALEEGVMSWSPKETYETDLAVLQWVEE 507
QY 1097 FGECVIALQASVIKKFLOGFMAP----- 1119
Db 508 LEQRVVLSDLQI-----RGWCTPTDSTREDITYCHELPDSDPEDIPWGRGREGTVPQK 562
QY 1120 -----KQKRRKLQSEDSAKTEBEVDEKQKVE-----BAKVA 1150
Db 563 NNNPLDLAVMLAVLEQNVVERRYLREPLWAAHEVVVVKALLSTPNGAPDGTSTEISYEIT 622
QY 1151 SALEKWKTAIREAQTFSRMHVLLGMLDACIKWDMSAENARCKVCPKKGDDKLILCDECN 1210
Db 623 PRVRVWRQTLERCRAAQVCLCMGQLERSIAEKSQVNVKTVCLVCRKGDNDLFLLCDCGD 682
QY 1211 KAFHLFLCLRPALYEPDGEWQCPACOPATARRNSGRNRYTESASEDSEDEDEEBE 1270
Db 683 RCGHIYCHRPKMEAVPEGDWDFCAVCL----- 708
QY 1271 EEEEEEDYEAVAGLRPRKTIKSHVIPPAAARSGRRPKKPHSTRSQPKAP--P 1325


```

RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426065; PubMed=12537568;
RA Celniker S.E., Wheeler D.A., Kronmiller B., Carlson J.W., Halpern A.,
RA Patel S., Adams M., Champagne M., Dugan S.P., Frise E., Hodgson A.,
RA George R.A., Hoskins R.A., Laverty T., Muzny D.W., Nelson C.R.,
RA Paclele J.M., Park S., Pfeiffer B.D., Richards S., Sodergren B.J.,
RA Svirskas R., Tabor P.B., Wan K., Stapleton M., Sutton G.G., Venter C.,
RA Weinstein G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.;
RT "Finishing a whole-genome shotgun. Release 3 of the Drosophila
RT melanogaster euchromatic genome sequence.";
RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426070; PubMed=12537573;
RA Kaminker J.S., Bergman C.M., Kronmiller B., Carlson J., Svirskas R.,
RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,
RA Ashburner M., Celniker S.E.,
RT "The transposable elements of the Drosophila melanogaster euchromatin:
RT a genomics perspective.";
RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426069; PubMed=12537572;
RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
RA Bettencourt B.R., Celniker S.E., de Grey A.D., Drysdale R.A.,
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA Lewis S.E.;
RT "Annotation of the Drosophila melanogaster euchromatic genome: a
RT systematic review.";
RL Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).
RN [5]
RP SEQUENCE FROM N.A.
RX FlyBase;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE FROM N.A.
RX FlyBase;
RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; A5003808; AAM70936.1; -.
DR HSP; P11362; 1EVT.
DR GO; GO:0004685; F:calcium- and calmodulin-dependent protein k. . .; IDA.
DR GO; GO:0004687; F:myosin-light-chain kinase activity; IDA.
DR InterPro; IPR003961; FN.III.
DR InterPro; IPR008957; FN.III-like.
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003598; IG.c2.
DR InterPro; IPR003006; IG.MHC.
DR InterPro; IPR011009; Kinase like.
DR InterPro; IPR007119; Prot.kinase.
DR InterPro; IPR002290; Ser_thr_kinase.
DR Pfam; PF00041; fn3; 1.
DR Pfam; PF00047; ig; 5.
DR Pfam; PF00069; Pkinase; 1.
DR ProDom; PD000001; Prot_kinase; 1.
DR SMART; SM00060; FN3; 2.
DR SMART; SM00408; IGC2; 12.
DR SMART; SM00220; S.TKC; 1.
DR PROSITE; PS00853; FN3; 2.
DR PROSITE; PS00835; IG.LIKE; 27.
DR PROSITE; PS00290; IG.MHC; UNKNOWN_1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; UNKNOWN_1.
DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; UNKNOWN_1.
DR ATP-binding; Kinase; Transferase.
KW
SQ SEQUENCE 9270 AA; 1036760 MW; 0DD82D040B27C2AD CRC64;
Query Match 4.4%; Score 352.5; DB 2; Length 9270;
Best Local Similarity 19.5%; Fred. No. 1.7e-05;
Matches 326; Conservative 260; Mismatches 559; Indels 523; Gaps 76;
QY 55 TGSOLLTHKEAWEERQEAELLKERFPAYWEKLVLEMVHHNTASLEKLVDTAWLEIMTKY 114
DB 3653 TVSEKITDEKAQESQKEB---VKSEAKPKKAKVLE---KKSIEEKLDDK--EKQTES 3704
QY 115 AVGECD----FEVGKEK---LKVIVKHIPLEKVVBEATEKKSD 153
DB 3705 AIDEKSQAQAEVSEIVSEKITDEKAQESQKEVKQSEAKPKKAKVLEKKSIEEKLDDKE 3764
QY 154 GACDSPSSDKENSSQIAOHQKKEIVVEDEGRRESINDRARRSPRKPTLSLKG--RRK 211
DB 3765 TQTDSAIDEKSQAQAEVSEIVSEKITDEKAQESQKEVKD--SEAKPKKAKVLEKKSIEE 3823
QY 212 WAPPKFLPHKYDVVKLQNDKIISNVPADSLI--RTERPPNKEIVRYFIRHNALRAGTGEN 269
DB 3824 LEDKKEKQTESAIDEKSQAQAEVSEIVSEKITDEKAQESQKEV-----KGSE 3870
QY 270 APWVVEDLVKYSLPKSFDFLLDPYKYMTLNPSTKRNVTGSPDRPKSKSK----- 322
DB 3871 AKPKKAKVLEKKSIEEKLDD-----KKEKQTESAIDEKSQAQAEVSEIVSE 3916
QY 323 --TDNSSLSPL-----NPKLWCHVHLKSLSGSLPKVKNSKNSKSPBEHLEMMKM 372
DB 3917 KITDEKAQESQAEVSEIVSEKITDEKAQESQKEV---NKEKQTESAIDEKSQA 3972
QY 373 MSPNKLHNFHPKPKGPPA--KPKGKHSD--KPLKAKGRSGKIINGOKSTGNSKSPKGLKT 430
DB 3973 AEVSEIVSEKITDEKAQESQKEVKDSEAKPKKAKVLEKKSIEE-----LED 4021
QY 431 PRTKMKQMTLLDMAKGTQMTAPRNSGGTPTRTSSKPHKHLPPAALHLIAYYKENDRED 490
DB 4022 KKEKQTESAIDEKSQAQAEVSEIVSENI--TDEKAQESQ-----KEYKDSA 4066
QY 491 KESALSCVISKTARLLSSDDRARLPEELRSIVQKEYE-----LLEHKRWASMSR- 540
DB 4067 K-----PKKAVL-----EKKSIEEKLDDKKTQTDSDAIDEKSQAQAEVSEI 4108
QY 541 -----EORKEYLK-----KKEELKKL--KEKAKERREKEMLER--EKQKY 580
DB 4109 VSEKITDEKAQESQAEVSEIVSEKITDEKAQESQKEVSEIVSEKITDEKAQESQKEVSEI 4168
QY 581 EDQELTGKRLPAFLVDTPEGLPNTLFGDVAMVVEFLSCYSGLLLPDQAQYPTAVSLMBA 640
DB 4169 EVSEIVSENI-----TDEKAQESQKEV-----KDSEAKPKKAKVLEK 4206
QY 641 LSADKGGFLYLNRLVILLQTLLOLLODETAEDYGEIGMKLSEIPLTLHS---VSEIV 696
DB 4207 KS-----IEEKLDDKKE---KQTESAIDEKSQAQAEVSEIV 4239
QY 697 RLCLRRSDVQEESEGSDTDNK-----DSAAFEDNEVQDE-----FLEKLETS 739
DB 4240 SEKITTEKAQ--ESQKEVKDSKAKPKKAKVLEKKSIEEKLDDKKTQTDSDAIDEKSQA 4298
QY 740 EFFELTSEKJQILTALCHIRILMTVSVQDHMETROQMSAELWKRVLAVKEENDKKRABK 799
DB 4299 EVSEIVSEK-----ITDEKAQESQ-----KEEVKDSSEAKP 4328
QY 800 QKKKEMEAKNKENGKVENGLGKTDRKRIVKPEPQVDTAEADM--ISAVKSRLL--AIQA 856
DB 4329 KKAKVLEKKSIEEKLDD-----KKEKQTESAIDEKSQAQAEVSEIVSEKITDEKAQ 4380
QY 857 KKEIREIQRERMKVK---LEROA--BEERIR--KHKAAAEKAFQEGIAKALVMRRTPIGTD 910
DB 4381 SOKKEVKGSEAKPKKAKVLEKKSIEEKLDDKKEKQTESAIDEKSQA----- 4428
QY 911 RNHNRYLWLFSDVPGLFTEKGWVHDSIDYRFRNHCHDHTVSGDEDYCPRSKANKLGKNS 970
DB 4429 -----EVSEIVSEK-----ITDEKAQESQKEEV-----KDSEAKPKKAKVLEK 4468

```


"The transposable elements of the Drosophila melanogaster euchromatin: a genomics perspective.";

Genome Biol. 3:RESEARCH0084-RESEARCH0084 (2002).

[4]

SEQUENCE FROM N.A.

MDLINS=2242669; PubMed=12537572;

Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S., Hradecky P., Huang Y., Kaminker J.S., Milburn G.H., Prochnik S.E., Smith C.D., Tupy J.L., Whitfield J.S., Bayraktaroglu L., Berman B.P., Bettencourt B.R., Celisner S.E., de Grey A.D., Drysdale R.A., Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q., Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M., Lewis S.E.;

"Annotation of the Drosophila melanogaster euchromatic genome: a systematic review.";

Genome Biol. 3:RESEARCH0083-RESEARCH0083 (2002).

[5]

SEQUENCE FROM N.A.

FlyBase;

Submitted (SEP-2002) to the EMBL/GenBank/DBSJ databases.

[6]

SEQUENCE FROM N.A.

FlyBase;

Submitted (MAR-2004) to the EMBL/GenBank/DBSJ databases.

[7]

EMBL; AE003669; AAF53983.2; --

HSP; Q9UIG0; IF62.

FlyBase; FBgn026577; CG8677.

GO; GO:0003677; P:DNA binding; IEA.

GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.

InterPro; IPR004022; DDT.

InterPro; IPR001965; Znf_PHD.

Pfam; PF00628; PHD; 1.

SMART; SM00249; PHD; 1.

PROSITE; PS50827; DDT; 1.

PROSITE; PS01359; ZF_PHD_1; UNKNOWN_1.

PROSITE; PS50016; ZF_PHD_2; 1.

SEQUENCE 2759 AA; 303551 MW; 149C31F6F4E7F47 CRC64;

Query Match 4.3%; Score 342; DB 2; Length 2759;

Best Local Similarity 19.0%; Pred. No. 1.1e-05;

Matches 265; Conservative 182; Mismatches 442; Indels 504; Gaps 55;

61 THKEAWEE-----EQVAELLKEEPAWTEKLVLEWHHTASLEKLVDTAMLEIMTKYAV 116

603 TESEAIEEGRNRTSVLQNNNSDSSSIQTQDQHKVPADLNN-----TKKPV 652

117 GECDPEVGKEMLVKVIKHPLEKVEEATEKSDGCDSPSSDKENSQIAQDQKK 176

653 TEKCVLESG---TVEDKLVINH-----QISTNALDTACDEKLS-CETESVPVNHDSK 702

177 ETVVKED-EGRRSINDRARRSPRLPTSLKKGKRWAPPKPLPKHYDKVLQNEDKIISN 235

703 MKSEQETATSAKESVSESS-----PTD---GEDETAKSKLIDPYAGKTNDKGIILNK 752

236 V-PADSLRT--ERPKNKEIVRYFTRNALRAGTGENAPWVVEDELVKKYSLPKSPDFL 292

753 IDKAEVTKLIQDDPPKDKISSETNTTCLINISPSAEHRISE-----796

293 LDPKYMTLNPSTKTK-----NTGSPDRKPSKSKTNDNSLSPNPLKLVCHVHKKS 345

797 ----KVQTEPSTQNLFEDNGKSGVDGKSRSTGAVEEISKTSTL-----LNRKR 845

346 LSGSPKVKNSKNSKSPHELEEMMKMSPNKLHTNFHPKKGPPAKPKGKSHDPLKAK 405

846 LNDQSALRNS--TSESEVHEEE-----PQDDDTLDDLVGGKRIKWR 887

406 GRSKILNGQKSTGNSKSPKGLKTPKTKQKMTLLDMAKGTQKWRAPRNSGGTPRTSS 465

888 PK-----TTNAESRRK-VEAQKTQIEET-----SSGDEDDARIRKCI 925

466 KPHKLPALALHLIAYYKKNKDKRNSALSCVISKTARLLSSDRLPEELRSLVKQR 525

926 APHTTKTPTLEEII-----ERLKKSI-----ENDLPKTKTEKVMEM 962

QY 526 YELLEHKRW-----SMSBQRKEYLKKKEELKKLKEKAKERREKEMLERLEKQKRY 580

Db 963 PTLMQKTREATPTIISLSPOKPTITKPLKNLLTLQLRQ---ESEBETIPR---KRT 1015

QY 581 EDQELTGNLPAFLVDTPEGL-----PNTLFGDVAMVVEFLSCYSGLLLPDAQVP 631

Db 1016 NSETLV-PAIPASNVLCQFDERHRRKRSSDANEAFSKESSPTE-----VP 1060

QY 632 ITAVSLMALSDAGKGFYLNRLVILLQTLQLQDEIAEDYDGLGMLKSEIPLTLHS 691

Db 1061 PSNVS--EKLXNN-----EQDIOEEV-----EDFLAM-S 1087

QY 692 VSELVRLCLRRSDVOESEE-----GSDDTDNKKDSAAAFEDNEVDQFLFKLETS 739

Db 1088 VKDSLRSADQSPVPEGSARRSGRRGPAVHSELQPKRTGGARDKMQPEVNAELKQ- 1146

QY 740 EFFELTSEKLIQLTALCHRILMTYSVQDHMETROOMSAELWKEKRLAVLKEENDKKRAEK 799

Db 1147 ---ESEDEKI-----STKIKSEAKDDPAPE-----SPENRRKKPEEK 1180

QY 800 QKRKEMEAKNKGKVENGLGKTRKRIKVFPEPOVDT-----EADMTISAVKSRRLAIQ 855

Db 1181 PIKEE-----PNEEPKPKVGRGPRKKR-----SVDTNIIETNDSETPPVRSRIAAQ 1230

QY 856 AKEREIOEREMVKLERQAEERIRKHAAAEKAFQEGIAKAKLVLMRRTPIGTDNRNHR 915

Db 1231 KIKSEARRKQBEVALRTMKQE--LKKKKAKKEA-----1263

QY 916 YWLSDEVPGLFIEKGWVHDSIDYRFNHCHKQHTV---SGDEDYCPRSKANKLGNKSNW 972

Db 1264 -----DPTVLEPSGEE-----1274

QY 973 TOHGTATVAVETTPKO-----GONLWFLCDSQKELDELLNCLHPQGIRESOLKLEKR 1028

Db 1275 -SESEASEAEARNKKKCPGDGW--SSDSEQ-----PESEEEEEEPH 1319

QY 1029 YQDIHSIHLARKPNLGLKSCDGNQELNLFRLSDLIEVATRLQKGLGYVEETSFEARV 1088

Db 1320 YE-----TDPGSPL---PRSD-----HEFSP- 1338

QY 1089 ISLEKLDPGSCVIALQASVIKFLQGPMAKQKRKLQSEDSAKTEVEVDEKKNVSEAK 1148

Db 1339 ---SELEDESQV-----PMKARTVYKENADDEEDAE-----1371

QY 1149 VASALEKWKTAIREAQTPSRMHVLLGLMDACIKWMSAENARCKVCPKKGEDDKLILCD- 1207

Db 1372 -----ACQCKGSDHPWILLCDT 1390

QY 1208 -ECKNAPHLCLRPALYVDPDGEWQCPACQ-----PATARNRSRGRNYTE 1251

Db 1391 PTCNGYHCSCLSPVLYIPEGDWHCPQCQEQELIALEROLEQYDTLVAQKQERILAE 1450

QY 1252 ESASEDSEDDSS-----DESEEEEEEEEDVEAGLRPRKTIIGKHSIVIPAARS 1305

Db 1451 EQAEREROEAEATLAANDENFKSEKEDDDDDMAVGKAEKVKRRRGDGRINRAAKR 1510

QY 1306 G--RRPGKKPHSTR------SOPKAPPVDDAEVDELVLQTKRSSRRQSLE 1348

Db 1511 GTRRRRGNSOSSHRKSLGSGSRSGSDSSDSTSFSDSDDEDFIYKLRKRQINVSFR 1570

QY 1349 LQKCEILHKKIVK 1361

Db 1571 LNEYDDLLINSALK 1583

RESULT 36

Q9V602

ID Q9V602 PRELIMINARY; PRT; 3080 AA.

AC Q9V602;

DT 01-MAY-2000 (Tremblrel. 13, Created)

DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)

DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)

DE CG10897-PA.
 GN Name=tou; ORFNames=CG10897;
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20196066; PubMed=10731132; DOI=10.1126/science.287.5461.2185;
 RA Adams M.D., Celnikier S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
 RA Abril J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrieria S., Fleischmann W.,
 RA Fostler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glöck A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Haddix N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milghina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D., Scheefer F., Shen H.,
 RA Shue B.C., Siden-Kianos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.Y., Wasarman D.A., Weinstock G.M., Weissenbach J.,
 RA Williams S.M., Woodgett J., Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
 RA Yeh R.F., Zaveri J.S., Zhang M., Zhang G., Zhao Q., Zhou L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426065; PubMed=12537568;
 RA Celnikier S.E., Wheeler D.A., Krommiller B., Carlson J.W., Halpern A.,
 RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,
 RA George R.A., Hoskins R.A., Laverty T., Muzny D.M., Nelson C.R.,
 RA Pacleb J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,
 RA Svirskas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
 RA Weinstock G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.;
 RT "Finishing a whole-genome shotgun: Release 3 of the *Drosophila*
 RT melanogaster euchromatic genome sequence.";
 RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426070; PubMed=12537573;
 RA Kaminker J.S., Bergman C.M., Krommiller B., Carlson J., Svirskas R.,
 RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,
 RA Ashburner M., Celnikier S.E.;
 RT "The transposable elements of the *Drosophila melanogaster* euchromatin:
 RT a Genomes perspective.";
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).
 RN [4]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426069; PubMed=12537572;
 RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,

RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,
 RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
 RA Battencourt B.R., Celnikier S.E., de Grey A.D., Drysdale R.A.,
 RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
 RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
 RA Lewis S.E.;
 RT "Annotation of the *Drosophila melanogaster* euchromatic genome: a
 RT systematic review.";
 RL Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).
 RN [5]
 RP SEQUENCE FROM N.A.
 RG FlyBase;
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
 RN [6]
 RP SEQUENCE FROM N.A.
 RG FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AE003825; AAF58638.2; -.
 DR HSSP; Q9UIG0; 1F62.
 DR FlyBase; FBgn0033636; tou.
 DR GO; GO:003677; F:DNA binding; IEA.
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
 DR InterPro; IPR000194; ATPase a/bcentre.
 DR InterPro; IPR001487; Bromodomain.
 DR InterPro; IPR004022; DDT.
 DR InterPro; IPR001739; Methyl-CpG_bind.
 DR InterPro; IPR006162; Ppantne_S.
 DR InterPro; IPR001965; Znf_PHD.
 DR InterPro; IPR001876; Znf_RangGDP.
 DR Pfam; PF00439; Bromodomain; 1.
 DR Pfam; PF02791; DDT; 1.
 DR Pfam; PF01429; MBD; 1.
 DR Pfam; PF00628; PHD; 2.
 DR PRINTS; PR00503; BROMODOMAIN.
 DR SMART; SM00297; BROMO; 1.
 DR SMART; SM00249; PHD; 2.
 DR SMART; SM00547; Znf_RBZ; 1.
 DR PROSITE; PS00152; ATPASE ALPHA BETA; UNKNOWN_1.
 DR PROSITE; PS50014; BROMODOMAIN_2; 1.
 DR PROSITE; PS50827; DDT; 1.
 DR PROSITE; PS00012; PHOSPHOPANTHETINE; UNKNOWN_1.
 DR PROSITE; PS01359; ZF_PHD 1; UNKNOWN_1.
 DR PROSITE; PS50016; ZF_PHD_2; 2.
 SQ SEQUENCE 3080 AA; -330285 MW; 32A4D6066F071AB3 CRC64;
 Query Match 4.3%; Score 342; DB 2; Length 3080;
 Best Local Similarity 15.5%; Pred. No. 1.3e-05;
 Matches 345; Conservative 306; Mismatches 531; Indels 1040; Gaps 83;
 QY 69 EQVAVELLKEEFPAWYKLV-----LEMVHNTASLEKLVDTAWLEIM 111
 DB 1003 EQQSNLSRENFSAIRAIVGSFLQAPPPYANDGEYIRMTDEDVA--KRLED---LKVF 1057
 QY 112 TKYAVGECDEFGVGEK--MLKXIVIKIHPLEKDEATEKKSCACDPSDSSDSENSQ1 169
 DB 1058 TRQILNVQRIEIAKQQQAMDAKIQKEELARNKEKARQEK-NSKLQQRKDELKNQ 1116
 QY 170 AODHOKKE----TVVKEDEGRRESINDRARRSPKLPSTLKKGKRWAPPKPLPHKYD-- 223
 DB 1117 AVEERERRRQHSILRMLERK-FEDREKKKHQLVLDRLRLRRMA-----ERKRDAE 1170
 QY 224 ----VKLQNEDKIISN---VPADSLIRTERPNKE-----IVRYEIR----- 258
 DB 1171 ILQILRRPNDESEMPQELVIPELDRINAGRLPGQAMDLLMVFELHNFGETLGFDMBSL 1230
 QY 259 -----HNALRAGTGENAPWVVEDEL-----VKYSLP----- 285
 DB 1231 PSLQNLHDALMSDSDNADA-----EELLSVWTHLLVCAIEDPGVNPGRHTLLGOSLNA 1286
 QY 286 ----SKFSDFLLDPKYMTLNPSKTKRKNVSGPDRKPSKK----- 320
 DB 1287 DITNSNVSE-ILRIYLYATATGEVQRMGHITVDRERRRVPDHHQLDSDTTTTHSHSVKNQ 1345


```
QY 321 -----SKTNSLS-----PLNP-----KLWCHV----- 340
Db 1346 EYKJLHENDTWKLSKDRPFPVALNPTKQAQMLAHLCNDLLMNAVLQIDGSLTCA 1405
QY 341 HLKSLSGSPLKVKNSK-----NSKSPHEHEMMKMSPNKLH----- 379
Db 1406 QMRKEKYMTDMKVKYKALHMRKARIEAYERAQAREAAQALMAQOKLDAERLKAEBEA 1465
QY 380 -----TNFHIPKGGPP-----AKPGHSDKPLKAG----- 406
Db 1466 KAAAAEAAAAAGTDEATKGGSPNGKEPBDGQNEGAKEPQQQQQPMVEYDVVDEA 1525
QY 407 -----RSKGILNGOKSTGNSPKK-----GLKTPKTKMONTLLDMAKGTQKMT 451
Db 1526 SLVSPAKTII-----QTDNSLTPSKQDMPTPTYQINGSSTPTTS-----GVTGDMNVL 1574
QY 452 RAPNSGGTPTSSKPKHLPAPALHL-----IAYKENKORE-----DK--RS 493
Db 1575 LOAKKSGARNSINDEHHHDVSIIDDDLSLDSEITNVEEDEDNRLSADELOKLDKLVRA 1634
QY 494 ALSC--VISTARLLSS-----EDRA-----RLPEE-----LRSIVQRYELLEHKKRWASM 538
Db 1635 SLNCKEALKSTNLQRAACFGQDRFWRRYWKLPKAGGIFIEALESQNDICDYHEALEAM 1694
QY 539 SEORKEYLKKKEELKKLKEKAKERRE--KEMLERLE-----KQRYEYO 583
Db 1695 DDKKANDKENSENEKVAEESSEQPMVEDESIITKLEDGPASDVGMPESNQONAHQDE 1754
QY 584 ELTGKGLPAFLV-----DTPGLNLTFGDVAMVVEFLSCYSGLLLPDAQVP--- 631
Db 1755 EDDDDVTEINKVEPIVDLGDODDDAAPP-----LPIEPPE 1794
QY 632 ITAVSLMEALSADGGLVNLRVLLQTLQDEIAEDYBELGMLKLSIPLTLHS 691
Db 1795 IKVSEMELMGPP-----TTMIST-----KTDF-EAEIKIPSPGILMP 1833
QY 692 VSLVRLCLRRSDVQSESGSDTDNKSAA-----FEDNEVDEFEKL----- 736
Db 1834 PT-----LNNNTNNNNNGSDCKLETGLGLGQQQONFSQSIVKTEDVKKEDDCIIVS 1889
QY 737 -----ETSBEF-----ELTSEE-----KLOILTA--LCHRIILWTSVDOH--- 769
Db 1890 TSSVDDTPKWFSLVRREVPLISELPAEBEGVVGQELQISVANQNSAQL-----OLQHPWD 1946
QY 770 -----METROQMAELMKE---RLAVLKEENDKRAKQKQKEMEAKNKENGKV 815
Db 1947 LINNQYYSIPMDECKVDTSKLGNECIFSLSGLDEKQMLAKVEEYKAHKVESK----- 1999
QY 816 ENGLG-----KTORKRIVKPEPOVDTE-----AEDMISAVKSRLLA--IOAKKEREI 862
Db 2000 -NGLGSPHRHETKODDEQAKLKDKEIDTEMETDADDLAGKEKFPRLRSVPPDTGGGV 2058
QY 863 QE-REMKVKLEQAEERIRKHAABAKAFQEGIAKAKLVMRT-----PIGTDNRNR 915
Db 2059 SEGTUVKPKIEURLDE-----ALSQAYYHNIAANSLSVVQTYIPIDIPLSMTPE 2110
QY 916 YML-----FSDEVPGLFIEKGWVHSDIDYRFNHHCKDHTVSGDEDYCPRSKKAANLCK 967
Db 2111 HRLLEQVKLAGPFRVHVYV----- 2131
QY 968 NASMNTQHTATVAVETTPKQGNLWFLCDSQKELBELLNCLHPQGIRESOLKERLKB 1027
Db 2132 -----PRQRYGWMQDDEQKRLQLLTKLTNPGLRERELQNLQR 2171
QY 1028 RYQDIITHSIHLAKPNLGLKSCDGNQELANF--LRSDL-----IEVATPL 1070
Db 2172 -----FLGLEQPLG-----VNYKLSIDIDFPEFLMPDKKGDWPNPKVAKRV 2212
QY 1071 QKGGLGYBETSFEARV-----ISLEKLD-----FGE 1099
Db 2213 E---LALIQLSELEKDVASASMLKNWQLPNRVSELTDSQEDVTEEDFVSIIPMIRE 2269
QY 1100 CVIALQASVKKFLQGFMAPK-----QKRRKLOSE 1129
```

RESULT 37

```
QY5PR8
ID Q75PR8 PRELIMINARY; PRT; 1637 AA.
AC Q75PR8;
DT 05-JUL-2004 (TREMELrel. 27, Created)
DT 05-JUL-2004 (TREMELrel. 27, Last sequence update)
DT 05-JUL-2004 (TREMELrel. 27, Last annotation update)
DE Unichrom.
OS Hemientrotus pulcherrimus (Sea urchin).
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC Echinoidea; Euechinoidea; Echinacea; Echinoidea; Strongylocentrotidae;
-OC Hemientrotus.
OX NCBI_TaxID=7650;
RN [1]
RP SEQUENCE FROM N.A.
RA Moritani K., Tagashira H., Shimotori T., Tanaka S., Takata K.,
RA Mitsuunaga-Nakatsubo K., Shimada H., Sakamoto N., Yamamoto T.,
```


Qy	1186	-----AENAR-----	1190
Db	2510	QAPLSIKINLKAQAQOCLLTKYTPILLTKSPASSSHPTSDSDSGKRTKKSGG	2569
Qy	1191	-----CKVCPKGGEDDKLILDCENKNAFLCLRPALVEYVDPGEW	1230
Db	2570	KRRBSANNTSSKYSYSLQNCQFCTSGENEDKLLLCDCGDKGYHTYCFKPKMDNIPGDW	2629
Qy	1231	QCPAC-QPATARN-----	1249
Db	2630	YCEYCNKATNERKCIYCGGHRFSPVGKMYCDLCPRAYHADCYIPPLLKVPGRKMYCHG	2689
Qy	1250	-----TEESASESDEDE-----	1283
Db	2690	CISRAPPPKRSAGCTSGSSKSRDRDRRESGSGSKRSNSKTPANEHMQQQOMPLAGG	2749
Qy	1284	-----LRLRPRTKIRG-----	1294
Db	2750	DSHHHHQPPSLNSSHDESNLSPAGSEVSGACGLNGRRSTRSGFQLNNASALCSP	2809
Qy	1295	KHSV-----TPPA-----	1302
Db	2810	AHSVVSATNYDDQHHANNSVDGSRFHAHLIPPSNNGTAALEEDVPGGANMFGVPEVYT	2869
Qy	1303	-----ARSGR-----	1307
Db	2870	PVAAGNFSAGLINOAPVQPMPPFANVVMSPRAVTPTRTPPTPTPTPTPTPTPTPTPLM	2929
Qy	1308	-----RPGKKPHS-----	1335
Db	2930	QASPTATALHVNAQSPFPQOQHAQLMTMPPSPPAIGVGTATNQWSPPINTHAQEAKEL	2989
Qy	1336	LQTKR-----SRRQSLLEKQEEILHKIVKRFSPFPREPTRDEADYDVITHPMDPQT	1392
Db	2990	KQEKKEKHATKMLKELAVCKTLGEMELHEDSWPFLLPVNTKQFTYRKIIKTPMDLST	3049
Qy	1393	VONKSCSGSYRSVQEFLLTMKQVFNAAEVN	1423
Db	3050	IKKKLQDLSYKTRDFCVDRQIRFDNCMFN	3080
<p>RESULT 39</p> <p>Q6Z178 PRELIMINARY; PRT; 617 AA.</p> <p>AC Q6Z178</p> <p>DT 05-JUL-2004 (TrEMBLrel. 27, Created)</p> <p>DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)</p> <p>DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)</p> <p>DE Putative DDT domain-containing protein.</p> <p>GN Name=OJ1008_F08.22;</p> <p>OS Oryza sativa (japonica cultivar-group).</p> <p>OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;</p> <p>OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;</p> <p>OC Ehrhartoideae; Oryzaceae; Oryza.</p> <p>OX NCBI_TaxID=39947;</p> <p>RN [1]</p> <p>RP SEQUENCE FROM N.A.</p> <p>RA Sasaki T., Matsumoto T., Yamamoto K.;</p> <p>RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.</p> <p>DR EMBL; AP004017; BADI5468.1; -.</p> <p>DR InterPro; IPR004022; DDT.</p> <p>DR Pfam; PF02791; DDT; 1</p> <p>DR SMART; SM00571; DDT; 1.</p> <p>DR PROSITE; PS50827; DDT; 1.</p> <p>SQ SEQUENCE 617 AA; 71727 MW; 4615520664AFDEC5 CRC64;</p>			
<p>Query Match 4.1%; Score 329; DB 2; Length 617;</p> <p>Best Local Similarity 18.3%; Pred. No. 6.9e-06;</p> <p>Matches 195; Conservative 115; Mismatches 236; Indels 522; Gaps 35;</p>			
Qy	3	PLGRKPPPLVNP---LPGEPPFTTPIHTQEAFTREYEREARLERYSERIWTCKSTGSSG	59

Query Match	4.0%	Score 323	DB 1	Length 1790
Best Local Similarity	20.2%	Pred. No. 4.1e-05		
Matches 274	Conservative 222	Mismatches 471	Indels 392	Gaps 58
17	PCEPFFPIPTQAFRTREYEAERLERYSERIWTCKSTGSSOLTKEAWE	---	---	EE 69
Db	719	PDEBPISKI	-----SFBEVKLQOCTKLKGETISLOT	-----ETSETHNTEKLIATLNEH 771
Qy	70	QEVAE	-----LLKEFPFAWYKLVLEVMHNTA-SLEKLV	-----DT 105
Db	772	KELDEKQIOLNSHSLKENFS	-----ILETELKNVRDSLDEMTQLDVLTKDKENQOT	825
Qy	106	AWLEIMTKYAVGESC	-----DFEVGKEMLKVKIVIKHPLEKV	-----DEATEKKSDDG 154
Db	826	ALLEVKSTIHQEDS	-----IKLEKLEITLSQKKKABDGINCKMKGOLFALSREMQAVEEN	--- 882
Qy	155	ACDSPSDKENSQIAQDHQKKTVEVKEDEGRRESINDRARRSPRKLPTSLKKGKRWAP	214	
Db	883	-CKNLQKEKDSN	---VNHQKTEKSLKEDIAAKIT	----- 913
Qy	215	PKFLPHKYDVKLQNEDKIISNVPADSLIRTERPPNKEIVRYFIHNNALRAGTGENAPVVV	274	
Db	914	-----EIKAINENLEMKIQCNLSKEKHSIKELVEYKSRFQS	----- 952	
Qy	275	EDELVKYSLSPKFSDFLLDPYKMTLNPSTKRNKTSQDPKPSKSKSTNNSLSPLNP	334	
Db	953	HDNLVAK-LTEKLKS-LANNYKMQAENESLIKAV	-----EESKNESSTQLSNLQN 1001	
Qy	335	KLVCHVHLKSLGSLPKVKNKSNK	-----SPHEHLEMMKMSPNKLTNPHFKPKGPP 390	
Db	1002	KI	-----DSMS	-----QKENFQIERGSIENIKELQKLTISDLE-QTKSEIISKDS 1047
Qy	391	AKPKGKSHDKPLKAGRSKGLNGOKSTGNSKSPKGLKTPKTKMKQMTLLDMAKGTQKM	450	
Db	1048	SKDEYESOISLKEKLET	-----ATTANDENVNKISLTKTKEELEAELAYKNL	--- 1097
Qy	451	TRAPNSGQTP-RTSSPKPHKLPPAALHLIAAYKENKDRDKGSALSVCISTARTLSS	508	
Db	1098	---KNELETKLETSEKALKEVNEEHL	-----KEEKIQLEKEA	-----TETKQQLNSL 1143
Qy	509	---EDRARLPBELSLVOKRYELLEHKRWAS	-----MSEQRKEYLKKKEEL 554	
Db	1144	RANLESLEKEHEDLAQALKYEEQIANKERQYNEEISQALDEITSTQENESIKKKNDEL	1203	
Qy	555	KKKLKEKAKERREKMLERLEKQKRYEDQELTGKNLPAFLVDTPPEGLPNTLFGDVAMVV	614	
Db	1204	-----EGEVKAMKSTSEQNLKKESEIDALNLQIKELKKQVETNEASLESIKS-V	1353	
Qy	615	EFLSCYSGLLLPDAQYPTITAVSLME	-----ALSADKGGFLYLNRLVILLOTLLOTLLODEI 671	
Db	1254	ESETVKIKELQDECNFKEVSELEDKLKASEDKNS-KYLE	-----LQKES 1298	
Qy	672	ADYQELGKMLSEPLTLHVSVELVRLRRSDVQSESGSDTDNDKDSAAFEDNEVQDE	731	
Db	1299	EKIKEELDAAKTTELKIQLEKITNL	-----SKAKEKSELSRLKKTSS-BERKNAEE 1349	
Qy	732	FLEKLETS-EFFELTSEKLIQILALCHRLILMTYSVODHMETRQMSAELWKERLAVKE	790	
Db	1350	QLEKLNRIQIKNOAFERKILLNEGSGTITQEYS	-----EKINTLE	-----DELIKLQN 1399
Qy	791	ENDKKRAE-KQKRKEMAKNKENGKVENGLKTKRKRIKVFEPQVDEADIMISAVKSR	849	
Db	1400	ENELKAKEDNTRSELEKVSLSNDEL	-----LSEKQNTIK	-----SLQDEILSYDKK 1446
Qy	850	-----RLLAIOAKKERIQE	-----REMKVKLERQAE	-----BERTRKHKAARAKAFQE 893
Db	1447	ITRNDEKLISIERDNKRDLESILKEQLRAAQESKAKVBEGLKLEESSEKKAEL	----- 1501	
Qy	894	GTAKAKLVWRTPIGTDNRNHRNRYMLFSDVEVPLFIEKGWVHDSIDYRFNHHCQDHTVSGD	953	
Db	1502	---KSKEMMKLLESTIESNETE	-----LKSMWE	-----TTRKS 1531


```
Db 1410 FREFNLSIHDITOLMELKR---ILYNLFLEYNPHNMNKKQ-----AELDKKFSMN 1460
Qy 1476 AVQSRDEDRRSREAE 1491
Db 1461 VHQILNYERIRMYE 1476

RESULT 50
Q6P5H2 PRELIMINARY; PRT; 1864 AA.
AC Q6P5H2
DT 05-JUL-2004 (Tremblrel. 27, Created)
DT 05-JUL-2004 (Tremblrel. 27, Last sequence update)
DT 25-OCT-2004 (Tremblrel. 28, Last annotation update)
DE Nestin.
GN Name=Nes;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Brain;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udén T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalek U., Smailus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Brain;
RA Strausberg R.;
RL Submitted (NOV-2003) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: Belongs to the intermediate filament family.
DR EMBL; BC062893; AAH62893.1; -
DR EMBL; BC060693; AAH60693.1; -
DR GO; GO:0005737; C:cytoplasm; IDA.
DR GO; GO:0007399; P:neurogenesis; IDA.
DR InterPro; IPR011000; Apolp_III_like.
DR InterPro; IPR001664; IF.
DR Pfam; PF00038; Filament; 1.
DR PROSITE; PS00226; IF; 1.
KW Intermediate filament.
SQ SEQUENCE 1864 AA; 207123 MW; B9DF21005D977983 CRC64;

Query Match 3.9%; Score 310.5; DB 2; Length 1864;
Best Local Similarity 19.5%; Pred. No. 0.00014;
Matches 314; Conservative 235; Mismatches 557; Indels 503; Gaps 73;

Qy 7 RKPFPVLNPLGSEPPFTTIPHTQE-AFRTREEYEARLERYSERIWTCKST-GSSQ----L 60
Db 160 RRP-----PAPAHASPIRAPEVEELARLLGEVVRGAVRDYQERVAHMESSLQARERLQ 214
Qy 61 THKEAWEERQEVALLKEBFPAPWYKLVLEVMVHHNTASLEKLVDTAWLEIMTKYAVGEBC 120
Db 215 AVRGARESRLEVOQLQADR-----DSLQERREALEQRLEGRWQD----- 253
```

```
Qy 121 DFEVGKMKLVKIVKIHPLEK--VDDEATEKKSGACDPSDSSDKENSSQIAQDHQKKT 178
Db 254 -----RLQATEKFQLAVEALEQEKQLSQIAQILEGGQQLAHLKMSLSL 298
Qy 179 VVKEDEGRRESINDRARRSPKLPISLKKGERK-----WAPPKF-- 217
Db 299 EVATYRTLLAEANSRLQTPGRSSQASLPDPKLLHLFLGIPEDQHLGSLVPLVSPTSFS 358
Qy 218 --LPHKYDVK-----LQNEDKIISNVPADSLIRTERPPNKIEIVRYFIRHNALRAG 265
Db 359 SPLPNTLETPTVAFLEKQFLKARTPTLASTIIPMSEAPYKNAEVRADQVPHSLQGG 418
Qy 266 TGENAPVWVEDELVKYSLSPKSFDFLLDPYKMTLNPSYKRNKGTSPDRKSKSKSDTN 325
Db 419 R-QQAP-----EPLWAEATVPSSTG-----VLPELEPPGGQPOPHFPD-----DP 457
Qy 326 SLSLSPINP-----KLMCHVHLKSLSSPLKVNKSNKNS 359
Db 458 TSLAPPLNPHHSILEAKDRESSESRVSIQBEQGQIW---ELVKKEAAATEVKVENSQAQ 514
Qy 360 KSPPEHL-----EEMKQW-----SPNKLHTNFHIPKKGPPAK 392
Db 515 EIQESGLDTEETQDSQGPLOMETLEALGDEPLMSLKTQNHETPGKENCNSSIENSSTVK 574
Qy 393 KPGKHSKPLKA-----KGRSGILNGQKSTGNS--KSPKGLKTPKTMMQMOTLL 441
Db 575 SPEKEKQTPKLSLEKNVNAEAKTLENGVLELSKPLGEESPRMEDQELMSPEHTLETVSFL 634
Qy 442 DNAQTQKMTAPRNSGGTPRTSSPKPKLPPAALHLIAYKENK-----DREDKRSAL 495
Db 635 --GKENQYVR-----SSEQNLE-----SLITFKESQYPLGGPAEDQ----- 672
Qy 496 SCVISTAKRLSSDDR--ARLPEE--LRSLVQKRYELLEHKR-----WASNSEQRKE 545
Db 673 -----MLRLVKEKQRPSPRSEEDQQAQPRLEKNQEPLEPEEADQVLERLIEKQEQ 727
Qy 546 YLKKKREE-----LKKLKEKAK-ERREKEMLERLEKQRY-----EDQELTGKN 589
Db 728 SLKSPSEEDQQAQPRLEKENQEPLEPEEADQVLERLIEKQESLKSPEEDDQ----- 782
Qy 590 LPAFLVDTPEGLPNTLFGDVANVVEFL---SCYGLLLPDAQYPTAVSL-----NEAL- 641
Db 783 --AFRLLEKENQEPLEPEEADQVLERLVEKESQESLSKSPSEEDQRTGKPLEKQESLR 840
Qy 642 SADKGGFLYNRLVILLQTLQTLQD--EIAEDYGELGMKLSLPTLTHSVSELVRLCL 700
Db 841 SLDE-----NOETIVLLESKNQRLRSLEVEEE-----ORIVKPLEKVSQVLSLESL 887
Qy 701 RRSVDQ-----ESEE-----GSDTDDNKDSAAAFEDNEVQDEF----- 732
Db 888 EXENVQSPRYLEDDHMIKSLLEDKTHEILGSLDRNGENFIPPENETQGSRLRPPEEDQ 947
Qy 733 -----LEKLETSEPFEL--TSSEKQILTALCHRILMTYSVDQHMETROOMAELWKERL 785
Db 948 RIVNHLEK--ESQEFRLSRPEAEERQV-----MVRSLG--ENHDPFLSVVKEEQM 995
Qy 786 AVLKEEND-----KRAEKQKREMAKKNKENGKLVENGLCKTDKKGRIVKFEPQVDTAEAD 841
Db 996 AESKLENSQDSRKSLEDESQTFGSLKENESLRSLAGQQE-----EQKLEQETQO 1049
Qy 842 MISAVKSRRLAIAQAK-EREI-----QEREMKVLERQAEERIRKHKAAAFKAFQEGI 895
Db 1050 PLRAVEDEQMTVPPEKVDPELPKPLRNDQEVRSLDKENQESVLSNKGME----- 1102
Qy 896 AKAKLVMRRTPTGTDNRNHNRYWLFDSFDEVPGLFIEKGWVHDSIDYRNFHHKDHHTVSGDE 955
Db 1103 -----TVKSETENIESLETVG-----EC----- 1121
Qy 956 YCPRSKKANLGNASNTQHGK-ATEVAVETTTTKQ-----GQNLWFLCDSQKEL 1004
Db 1122 -----LGRRSKVDTPQELWTEVTSSETIEPLEDETOEPLGCVDENQEVLTPLERES 1172
```

```
QY 1005 DELLN-----CLHPQIRESQLKERLE-----KRYODITHSIHLARKPNLGLKSCDG 1051
Db 1173 QELRSLGKNPTVSPGVEDSQCLEVEEGPEREQHOESURSLGEVEWELPG-----SG 1228
QY 1052 NOELNLFRLSLIEVATRLQKGLGYVETSFEARVI-----SLEKLDGFCECVIALQA 1106
Db 1229 SQQRW-----EDVDEGEG-QEASLGATGVETEDKAELHLRGOGGEBEKAVEGE---LLQD 1280
QY 1107 SVIKKFLQGFMAPKOKR-----RKLQSEDSAKTEEV 1137
Db 1281 AVGEAWSLGSSEPKQRPVPAEPLDLEGOEQGTGLEVPVQAQMPATEQDEDRAGAGEQ 1340
QY 1138 D--EEKWVEEAKVASALEKWKTAIREAQTSR---MHVLLGMLDACIKWDMASANARCK 1192
Db 1341 DSVETVLGLEAARAGLELEQEVVGLDPRHFAREEAIHPSLGEESVYKAKIDQGLE----- 1395
QY 1193 VCPKKGEDDKLILCDECNKAFHLFCLRPALYEVDPDGEWQCPACQPATARRNSRGRNYTEE 1252
Db 1396 ---EPGKEPK-----EAGALDSGIPELPKTSSETLECKGWEES-----GEGWGE 1437
QY 1253 SASSEDEDESEEEEEEEED-----YEVAGRLI-----RPRKTIRGHKSHVIPP- 1302
Db 1438 EASLETSDHEGSHAQPRPPKTEDEGLQAALTVPKLPKLEPCSPILITDAHELQPOAE 1497
QY 1303 --ARSGRRP-----GK---KPHSTRRSQPKA-----PPVDDAEVDEL 1334
Db 1498 GIOEAGWQPEAGTEALGRVEDEPFGRIPEGLQDWEEGREDSADEL 1546
```

Search completed: March 9, 2005, 14:24:02
Job time : 218.036 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 9, 2005, 13:06:37 ; Search time 110.963 Seconds
(without alignments)
5336.298 Million cell updates/sec

Title: US-10-702-148-29
Perfect score: 7985
Sequence: 1 MAPLLGRKPPFLVNLPGEE.....NCFMVLVNTQFCMALTDVT 1531

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 65 summaries

Database : A_Geneseq_16Dec04:*

1: geneseqp1980s:*

2: geneseqp1990s:*

3: geneseqp2000s:*

4: geneseqp2001s:*

5: geneseqp2002s:*

6: geneseqp2003as:*

7: geneseqp2003bs:*

8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7985	100.0	1531	2 AAW81173	Aaw81173 Human BAZ
2	7955	99.6	1527	2 Aaw81172	Aaw81172 Human BAZ
3	7658	95.9	1483	8 ADQ20910	Adq20910 Human sof
4	7656	95.9	1483	7 ADf69140	Adf69140 Human MP5
5	5307	66.5	1052	6 ABR41364	Abr41364 Human DIT
6	2877	36.0	572	4 AAm14838	Aam14838 Peptide #
7	2877	36.0	572	4 ABB33805	Abb33805 Peptide #
8	2877	36.0	572	4 AAm27265	Aam27265 Peptide #
9	2877	36.0	572	4 ABB28622	Abb28622 Peptide #
10	2877	36.0	572	4 ABB19248	Abb19248 Protein #
11	2877	36.0	572	4 AAm66978	Aam66978 Human bon
12	2877	36.0	572	4 AAm54571	Aam54571 Human bra
13	2877	36.0	572	4 ABG48640	Abg48640 Human liv
14	2877	36.0	572	4 AAm02563	Aam02563 Peptide #
15	2835	35.5	560	4 ABG36635	Abg36635 Human pep
16	2835	35.5	560	4 AAm18320	Aam18320 Peptide #
17	2835	35.5	560	4 ABB37354	Abb37354 Peptide #
18	2835	35.5	560	4 AAm30808	Aam30808 Peptide #
19	2835	35.5	560	4 ABB32101	Abb32101 Peptide #
20	2835	35.5	560	4 ABB22640	Abb22640 Protein #
21	2835	35.5	560	4 AAm70484	Aam70484 Human bon
22	2835	35.5	560	4 AAm58044	Aam58044 Human bra
23	2835	35.5	560	4 AAm05928	Aam05928 Peptide #
24	2835	35.5	560	5 ABG40123	Abg40123 Human pep
25	2305.5	28.9	513	4 AAm00760	Aam00760 Human bon

Abx41346 Human DIT
Aaw81169 Human BAZ
Adf69139 Human MP5
Adp12578 Protein e
Aay07734 Human htl
Aau16181 Human nov
Abu55250 Human nov
Abb58706 Drosophil
Aag03625 Mouse sec
Abm85416 Mouse pro
Aaw81171 Human BAZ
Abx64241 Angiogene
Adp54420 Human pro
Abm85417 Human pro
Aab95554 Human pro
Aam40239 Human pol
Aam42025 Human pol
Aaw81170 Human BAZ
Aab93746 Human pro
Adr09529 Human pro
Aag03100 Human sec
Abx52663 Protein s
Adx61762 Disease t
Abb58514 Drosophil
Aam14835 Peptide #
Abb33802 Peptide #
Aam27262 Peptide #
Abb28619 Peptide #
Abb19245 Protein #
Aam66975 Human bon
Aam54568 Human bra
Abg48637 Human liv
Aam02560 Peptide #
Abx36632 Human pep
Aam14836 Peptide #
Abb33803 Peptide #
Aam27263 Peptide #
Abb28620 Peptide #
Abb19246 Protein #
Aam66976 Human bon

ALIGNMENTS

RESULT 1
AAW81173
ID AAW81173 standard; protein; 1531 AA.

XX
AC AAW81173;
XX
DT 05-MAY-1999 (first entry)
XX
DE Human BAZI-beta protein #2.
XX
KW Transcriptional regulator; transcription; BAZI-alpha; bromodomain; BAZ;
KW atypical zinc finger; testis; human; tumour; BAZI-beta; BAZ2-alpha; drug;
KW BAZ2-beta; treatment; cancer; proliferative disorder; screening.
XX
OS Homo sapiens.
XX
PN WO9847920-A1.
XX
PD 29-OCT-1998.
XX
PF 17-APR-1998; 98WO-JP001783.
XX
PR 18-APR-1997; 97JP-00116570.
XX
PP 24-OCT-1997; 97JP-00310027.
XX
PI (CHUS) CHUGAI RES INST MOLECULAR MEDICINE INC.
XX
PI Jones MH;

XX WPI; 1998-583603/49.
 DR N-PSDB; AAV68409.
 XX
 PT Transcriptional regulator gene family containing bromodomain - may be
 expressed in testis tissue and is useful for treatment of cancer and
 other proliferative disorders.
 PT
 XX
 PS Claim 1; Page 145-157; 187pp; Japanese.
 XX
 CC This sequence represents the human BAZ1-beta protein, a member of a
 family of transcriptional regulator genes containing a bromodomain (BAZ,
 CC Bromodomain with Atypical Zinc finger) which are expressed specifically
 in testis tissue and also in certain tumour lines. Transgenic cells may
 CC be used for the preparation of the BAZ1-alpha, BAZ1-beta, BAZ2-alpha and
 BAZ2-beta proteins. These proteins can be used in the treatment of cancer
 CC and other proliferative disorders, and in screening of compounds for
 their binding ability to the expression products (e.g. for use as drugs
 CC by modulation of transcriptional regulation)
 XX
 SQ Sequence 1531 AA;

Query Match 100.0%; Score 7985; DB 2; Length 1531;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1531; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAPLLGRKPPFLVNPDPGEEFFPIPTQBAFRTRREYEARELRYSEIRWTCKSTGSSQL 60
 DB 1 MAPLLGRKPPFLVNPDPGEEFFPIPTQBAFRTRREYEARELRYSEIRWTCKSTGSSQL 60
 QY 61 THKEAWEEQVAAELKEEFPAWYEKLVLEWVHNTASLEKLVDTAWLEIMTKYAVGEEC 120
 DB 61 THKEAWEEQVAAELKEEFPAWYEKLVLEWVHNTASLEKLVDTAWLEIMTKYAVGEEC 120
 QY 121 DFEVGKEXMLKVIKIHPLEKVDDEATEKSDGACDSSPSDKENSSQIAQDHQKKEIVV 180
 DB 121 DFEVGKEXMLKVIKIHPLEKVDDEATEKSDGACDSSPSDKENSSQIAQDHQKKEIVV 180
 QY 181 KEDEGRRESINDRARRSPKLPSTSLKGERKWAPPKFLPHKYDVVKLQNEDKIIISNVPADS 240
 DB 181 KEDEGRRESINDRARRSPKLPSTSLKGERKWAPPKFLPHKYDVVKLQNEDKIIISNVPADS 240
 QY 241 LIRTERPNKEIVRYFIHNLALRAGTGENAPWVEDELVKYSLPSKESDFLLDPKYMT 300
 DB 241 LIRTERPNKEIVRYFIHNLALRAGTGENAPWVEDELVKYSLPSKESDFLLDPKYMT 300
 QY 301 LNPSTKRNTGSPDRKPSKSTDNSSLSPLNPKLMCHVHLKKSLSGSPKLVKNKSNK 360
 DB 301 LNPSTKRNTGSPDRKPSKSTDNSSLSPLNPKLMCHVHLKKSLSGSPKLVKNKSNK 360
 QY 361 SPEEHLEEMKMSPNKLTHTFHPKGPAPPAKPGKSHSDKPLKAGRSKGILNGOKSTGN 420
 DB 361 SPEEHLEEMKMSPNKLTHTFHPKGPAPPAKPGKSHSDKPLKAGRSKGILNGOKSTGN 420
 QY 421 SKSPKKGUKTPTKMKQMTLLDMAKGTOKMTRAPNSGGTPTRTSSKPHKLPALHLIA 480
 DB 421 SKSPKKGUKTPTKMKQMTLLDMAKGTOKMTRAPNSGGTPTRTSSKPHKLPALHLIA 480
 QY 481 YYENKDKREDKRSALSCVISKTARLLSSDDRARPPEELRSVQKRYELLEHKKRWASMS 540
 DB 481 YYENKDKREDKRSALSCVISKTARLLSSDDRARPPEELRSVQKRYELLEHKKRWASMS 540
 QY 541 EQRKEYLKKGEELKKLKEKAKEREKEMERLEKQKRYEDQELTGKPLPAFLVDTPPE 600
 DB 541 EQRKEYLKKGEELKKLKEKAKEREKEMERLEKQKRYEDQELTGKPLPAFLVDTPPE 600
 QY 601 GLPNTLFGDVAVVVEFLCYSGLLLPDAQYPTITAVSLMEALSADKGGFLYLNRLVILLQ 660
 DB 601 GLPNTLFGDVAVVVEFLCYSGLLLPDAQYPTITAVSLMEALSADKGGFLYLNRLVILLQ 660
 QY 661 TLLQTLQDEIAEDYGGELGMKLSIPLTLHVSSELVRLCLRRSDVQEESEGSDTDDNKDS 720
 DB 661 TLLQTLQDEIAEDYGGELGMKLSIPLTLHVSSELVRLCLRRSDVQEESEGSDTDDNKDS 720

QY 721 AAFEDNEVQDEFLEKLTSEFFELTSBEKLOILTALCHRILMTYSVODHMETROOMSAEL 780
 DB 721 AAFEDNEVQDEFLEKLTSEFFELTSBEKLOILTALCHRILMTYSVODHMETROOMSAEL 780
 QY 781 WKERLAVLKENDKKRAEKQKREMEAKNKENGKVENGLKGTDRKKRIVKPEPQVDTEAE 840
 DB 781 WKERLAVLKENDKKRAEKQKREMEAKNKENGKVENGLKGTDRKKRIVKPEPQVDTEAE 840
 QY 841 DMISAVKSRRLAIOAKKERIOEREMKVKLERQAEERIRKHAAAFAKAFQEGIAKAL 900
 DB 841 DMISAVKSRRLAIOAKKERIOEREMKVKLERQAEERIRKHAAAFAKAFQEGIAKAL 900
 QY 901 VMRRTPGTGNHNRNWLFSDEVGCLFIEKGWVHSDIDYRNNHCKOHTVSGDEDCPRS 960
 DB 901 VMRRTPGTGNHNRNWLFSDEVGCLFIEKGWVHSDIDYRNNHCKOHTVSGDEDCPRS 960
 QY 961 KKANLGNKSNMNTQGTATVAVETTPKQONLWFLCDSQKELDELINCLHPQGIRESQ 1020
 DB 961 KKANLGNKSNMNTQGTATVAVETTPKQONLWFLCDSQKELDELINCLHPQGIRESQ 1020
 QY 1021 LKERLEKRYQDI IHSIHLARKPNGLKSCDQGNQELLNPLRSDLIEVATRLQKGLGYVEE 1080
 DB 1021 LKERLEKRYQDI IHSIHLARKPNGLKSCDQGNQELLNPLRSDLIEVATRLQKGLGYVEE 1080
 QY 1081 TSEFEARVISLEKLDKFGECVIALQASVIRKFLQGFMAPKOKRRKLOSEDSAKTEEVDEE 1140
 DB 1081 TSEFEARVISLEKLDKFGECVIALQASVIRKFLQGFMAPKOKRRKLOSEDSAKTEEVDEE 1140
 QY 1141 KKMVEEAKVASALEKWKTAIREAQTFSRMHVLLGMLDACIKWDMSAENARCKVCPKKGED 1200
 DB 1141 KKMVEEAKVASALEKWKTAIREAQTFSRMHVLLGMLDACIKWDMSAENARCKVCPKKGED 1200
 QY 1201 DKLLILDECNKAFHLFCLRPALYEVDPGEWOCQACQATARNRSGRNYTTEESASEDS 1260
 DB 1201 DKLLILDECNKAFHLFCLRPALYEVDPGEWOCQACQATARNRSGRNYTTEESASEDS 1260
 QY 1261 DESDEEEEEEEDYEVAGLRLPRKTIKSHSVIPPAARSGRRPGKPHSTRSQ 1320
 DB 1261 DESDEEEEEEEDYEVAGLRLPRKTIKSHSVIPPAARSGRRPGKPHSTRSQ 1320
 QY 1321 PKAPPVDDAEVDLVLQTKSSRRQSLQKCEILHKIVKRYFSPFPREPVTDEADY 1380
 DB 1321 PKAPPVDDAEVDLVLQTKSSRRQSLQKCEILHKIVKRYFSPFPREPVTDEADY 1380
 QY 1381 YDVITHPMDFOTVQNKSCGYSYRVSQSFPLTDMKQVFNNAEYVNCRGSHVLSVMKTEOCL 1440
 DB 1381 YDVITHPMDFOTVQNKSCGYSYRVSQSFPLTDMKQVFNNAEYVNCRGSHVLSVMKTEOCL 1440
 QY 1441 VVLLAHKHLPGHPYVRRKPKFPDRLAEDGDESEPAVGQSDRDRSRREAEIQEWLQDTS 1500
 DB 1441 VVLLAHKHLPGHPYVRRKPKFPDRLAEDGDESEPAVGQSDRDRSRREAEIQEWLQDTS 1500
 QY 1501 LYSAKINSKDHNCFMMLVNTQFCMALTDVT 1531
 DB 1501 LYSAKINSKDHNCFMMLVNTQFCMALTDVT 1531

RESULT 2

AAW81172

ID AAW81172 standard; protein; 1527 AA.

XX AAW81172;

AC AAW81172;

XX 05-MAY-1999 (first entry)

XX Human BAZ1-beta protein #1.

XX

XX Transcriptional regulator; transcription; BAZ1-alpha; bromodomain; BAZ;

XX atypical zinc finger; testis; human; tumour; BAZ1-beta; BAZ2-alpha; drug;

XX BAZ2-beta; treatment; cancer; proliferative disorder; screening.

XX Homo sapiens.

OS


```

XX PN WO9847920-A1.
XX PD 29-OCT-1998.
XX PF 17-APR-1998; 98WO-JP001783.
XX PR 18-APR-1997; 97JP-00116570.
XX PR 24-OCT-1997; 97JP-00310027.
XX PA (CHUS) CHUGAI RES INST MOLECULAR MEDICINE INC.
XX PI Jones MH;
XX DR WPI; 1998-583603/49.
XX DR N-PSDB; AAV68408.
XX PT Transcriptional regulator gene family containing bromodomain - may be
XX PT expressed in testis tissue and is useful for treatment of cancer and
XX PS other proliferative disorders.
XX PS Claim 1; Page 125-137; 187pp; Japanese.
XX CC This sequence represents the human BAZ1-beta protein, a member of a
XX CC family of transcriptional regulator genes containing a bromodomain (BAZ,
XX CC Bromodomain with Atypical Zinc finger) which are expressed specifically
XX CC in testis tissue and also in certain tumour lines. Transgenic cells may
XX CC be used for the preparation of the BAZ1-alpha, BAZ1-beta, BAZ2-alpha and
XX CC BAZ2-beta proteins. These proteins can be used in the treatment of cancer
XX CC and other proliferative disorders, and in screening of compounds for
XX CC their binding ability to the expression products (e.g. for use as drugs
XX CC by modulation of transcriptional regulation)
XX SQ Sequence 1527 AA;

Query Match 99.6%; Score 7955; DB 2; Length 1527;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 1527; Conservative 0; Mismatches 0; Indels 4; Gaps 1;

QY 1 MAPLLGRKPPFLVNPVLPGEERPFTIPHTQEAAPTREVEEARLERYSERIWTCKSGSSQL 60
DB 1 MAPLLGRKPPFLVNPVLPGEERPFTIPHTQEAAPTREVEEARLERYSERIWTCKSGSSQL 60
QY 61 THKEAWEEQEAELLKEFPAPWYKLVLEMHVHNTASLEKLVDTAWLEIMTKYAVGSEC 120
DB 61 THKEAWEEQEAELLKEFPAPWYKLVLEMHVHNTASLEKLVDTAWLEIMTKYAVGSEC 120
QY 121 DFEVGERKMLKVIKVIHPLEKVDDEATEKSDGACDSPSSDKENSSQIAQDHQKKTWV 180
DB 121 DFEVGERKMLKVIKVIHPLEKVDDEATEKSDGACDSPSSDKENSSQIAQDHQKKTWV 180
QY 181 KEDEGRRESINDRARRSPKLTSLKGERKWAPPKELPHKYDVKLQNEDKIISNPADS 240
DB 181 KEDEGRRESINDRARRSPKLTSLKGERKWAPPKELPHKYDVKLQNEDKIISNPADS 240
QY 241 LIRTERPNKEIVRYFIRHNALRAGTGENAPWVEDELVKKYSLPKSFDFLLDPKYWT 300
DB 241 LIRTERPNKEIVRYFIRHNALRAGTGENAPWVEDELVKKYSLPKSFDFLLDPKYWT 300
QY 301 LNPSTKRKNTGSPDRKPKSKTNSLSPLNPKLWCHVHLKSLGSPPLKVKNSKNSK 360
DB 301 LNPSTKRKNTGSPDRKPKSKTNSLSPLNPKLWCHVHLKSLGSPPLKVKNSKNSK 360
QY 361 SPEEHLEEMKMSPNKLTNPHIPKGPAPKPKGKSDKPLKAGRSKGIILNGOKSTGN 420
DB 361 SPEEHLEEMKMSPNKLTNPHIPKGPAPKPKGKSDKPLKAGRSKGIILNGOKSTGN 420
QY 421 SKSPKGLKTPKTKNQMTLLDMAKGTQMTAPRNSGGTPTRTSSKPKHKLPPAALHLIA 480
DB 421 SKSPKGLKTPKTKNQMTLLDMAKGTQMTAPRNSGGTPTRTSSKPKHKLPPAALHLIA 480
QY 481 YYKENKREDKRSALSCVISKTARLLSSEDRARLPEELRSLVQKRYELLEHKRWASMS 540

```

RESULT 3
ADQ20910

```

DB 481 YYKENKREDKRSALSCVISKTARLLSSEDRARLPEELRSLVQKRYELLEHKRWASMS 540
QY 541 EORKEYLKKREBELKKLKEKAKERREKEMLERLEKQKRYEDQELTGKNLPAPRLVDTP 600
DB 541 EORKEYLKKREBELKKLKEKAKERREKEMLERLEKQKRYEDQELTGKNLPAPRLVDTP 600
QY 601 GLPNTLFGDVAMVVEFLSCYSGLLLPDAQYPTAVSLMEALSADKGGFLYLNRLVILIQ 660
DB 601 GLPNTLFGDVAMVVEFLSCYSGLLLPDAQYPTAVSLMEALSADKGGFLYLNRLVILIQ 660
QY 661 TLLOTLLODETAEDYGEIGMKLSEIPLTHSVSELVRLCLRSVDQSESGSDTDNKS 720
DB 658 -LLOTLLODETAEDYGEIGMKLSEIPLTHSVSELVRLCLRSVDQSESGSDTDNKS 716
QY 721 AAFEDNEVQDEFLEKLETSEFFELTSEEKLIQLTALCHRIILMTYSVQDHMETROQMSABL 780
DB 717 AAFEDNEVQDEFLEKLETSEFFELTSEEKLIQLTALCHRIILMTYSVQDHMETROQMSABL 776
QY 781 WKERLAVLKEENDKKRAEKQKREKMEAKVNGKVENGLGKTDRKKRIVKFPQVDTAE 840
DB 777 WKERLAVLKEENDKKRAEKQKREKMEAKVNGKVENGLGKTDRKKRIVKFPQVDTAE 836
QY 841 DMI SAVKSRRLIATQAKKERSEIOBERMKVKLERQAEERIRKHAAAEKAFQEGTAKKL 900
DB 837 DMI SAVKSRRLIATQAKKERSEIOBERMKVKLERQAEERIRKHAAAEKAFQEGTAKKL 896
QY 901 VMRRTPIGTDNRNHRNMYLFSDEVPGLFTEKGMVHDSIDYRFNHHCKDHTVSGDEYCPRS 960
DB 897 VMRRTPIGTDNRNHRNMYLFSDEVPGLFTEKGMVHDSIDYRFNHHCKDHTVSGDEYCPRS 956
QY 961 KKANLGKNASVNTQHTGTAETAVETTTPKQGNLWFLCDSQKELDELNCLHPQGISQ 1020
DB 957 KKANLGKNASVNTQHTGTAETAVETTTPKQGNLWFLCDSQKELDELNCLHPQGISQ 1016
QY 1021 LKERLEKRYODI IHSIHLARKPNGLKSCDGNQELNLFRLSDI IEVATRLQKGGYVEE 1080
DB 1017 LKERLEKRYODI IHSIHLARKPNGLKSCDGNQELNLFRLSDI IEVATRLQKGGYVEE 1076
QY 1081 TSEFEARVISLEKLDKDFGECVIALQASVIKKFLOGFMAPKQKRRKQSEDSAKTEVEDE 1140
DB 1077 TSEFEARVISLEKLDKDFGECVIALQASVIKKFLOGFMAPKQKRRKQSEDSAKTEVEDE 1136
QY 1141 KQWVEEAKVASALEKWKTAIREAQTFSRMHVLGLMDACIKWDMSAENARCKVCPKGGED 1200
DB 1137 KQWVEEAKVASALEKWKTAIREAQTFSRMHVLGLMDACIKWDMSAENARCKVCPKGGED 1196
QY 1201 DKLLCDECNKA FHLFCLRPALYEVDPDGEWQCPACQATARRNSGRNYTEESASEDS 1260
DB 1197 DKLLCDECNKA FHLFCLRPALYEVDPDGEWQCPACQATARRNSGRNYTEESASEDS 1256
QY 1261 DESDBEEBEEBEEBEEDEYEVAGLRRLRPKTIIRGHSVIPPAARSGRRPGKKPHSTRSQ 1320
DB 1257 DESDBEEBEEBEEBEEDEYEVAGLRRLRPKTIIRGHSVIPPAARSGRRPGKKPHSTRSQ 1316
QY 1321 PKAPPVDDAEVDELVLQTKRSSRQSLQKCEEILHKIKVYRFSWFPREPVRDEADEY 1380
DB 1317 PKAPPVDDAEVDELVLQTKRSSRQSLQKCEEILHKIKVYRFSWFPREPVRDEADEY 1376
QY 1381 YDVITHPMDFOTVQNKSCGYSYRVOEELTDMKQVFTNAEYVNCBGSVLSQWVTEOCL 1440
DB 1377 YDVITHPMDFOTVQNKSCGYSYRVOEELTDMKQVFTNAEYVNCBGSVLSQWVTEOCL 1436
QY 1441 VVLLHKLHPGHPYVRRKRRKFPDRLAEDGEGSEPEAVGQSDRDEDRRSREAEIQEWLQDTS 1500
DB 1437 VVLLHKLHPGHPYVRRKRRKFPDRLAEDGEGSEPEAVGQSDRDEDRRSREAEIQEWLQDTS 1496
QY 1501 LYSAKINSKDNCFMFLVNTQFCMALTDVT 1531
DB 1497 LYSAKINSKDNCFMFLVNTQFCMALTDVT 1527

```

ID ADQ20910 standard; protein; 1483 AA.
XX
AC ADQ20910;
XX
DT 26-AUG-2004 (first entry)
XX
DE Human soft tissue sarcoma-upregulated protein - SEQ ID 3730.
XX
KW soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human.
XX
OS Homo sapiens.
XX
PN WO2004048938-A2.
XX
PD 10-JUN-2004.
XX
PF 26-NOV-2003; 2003WO-US038193.
XX
PR 26-NOV-2002; 2002US-0429739P.
XX
PA (PROT-) PROTEIN DESIGN LABS INC.
XX
PI Aziz N, Ginsburg WM, Zlotnik A;
XX
WI; 2004-441208/41.
XX
XX Early detection of soft tissue sarcoma comprises determining expression
PT of a gene in a first soft tissue sample and a normal soft tissue sample
PT and comparing the gene expression, also useful in treating soft tissue
PT sarcoma.
XX
PS Example 2; SEQ ID NO 3730; 210pp; English.
XX
XX The invention relates to a novel method for detecting soft tissue sarcoma
CC which comprises obtaining a first soft tissue sample from an individual
CC and a normal soft tissue sample from the same or different individual,
CC determining the expression of a gene in both samples and comparing the
CC expression of the gene in both soft tissue samples, where a higher level
CC of protein expression in the first soft tissue sample indicates the
CC presence of soft tissue sarcoma. The method of the invention has
CC cytosstatic applications and may be useful for detecting soft tissue
CC sarcoma, possibly via gene therapy or vaccine production. The nucleic
CC acid sequences may be useful in diagnostic and screening applications.
CC The current sequence is that of a human soft tissue sarcoma-upregulated
CC protein of the invention. The current sequence is not shown within the
CC specification per se but was submitted in CD format by the inventor.
XX
SQ Sequence 1483 AA;

Query Match 95.9%; Score 7658; DB 8; Length 1483;
Best Local Similarity 98.9%; Pred. No. 0;
Matches 1473; Conservative 3; Mismatches 7; Indels 6; Gaps 2;

Qy 1 MAPLLGRKPPFLVNPDPGEEPPFPIPTQEAFTREEYEARELERYSERIWTCKTGSQ 60
Db 1 MAPLLGRKPPFLVNPDPGEEPLFIPTQEAFTREEYEARELERYSERIWTCKTGSQ 60
Qy 61 THKSAWEEQVAELLKEEFPAPWEKLVLEMVHNTASLEKLVDTAMLEIMTKYAVGEEC 120
Db 61 THKSAWEEQVAELLKEEFPAPWEKLVLEMVHNTASLEKLVDTAMLEIMTKYAVGEEC 120
Qy 121 DFEVGEKQKLVKVIHPLKVEDEATEKKSDGACDSDSPSDKENSQIAQDHQKETT 180
Db 121 DFEVGEKQKLVKVIHPLKVEDEATEKKSDGACDSDSPSDKENSQIAQDHQKETT 180
Qy 181 KEDEGRRESINDRARRSPKLPSTLSKKGERKWAFFKFLPHKYDVKLQNEDKIISNPADS 240
Db 181 KEDEGRRESINDRARRSPKLPSTLSKKGERKWAFFKFLPHKYDVKLQNEDKIISNPADS 240
Qy 241 LIRTERPNKKEIVYFIHNNALRAGTGENAPWVVEDELVKKYSLPSPKPSDFLLDPYKMT 300
Db 241 LIRTERPNKKEIVYFIHNNALRAGTGENAPWVVEDELVKKYSLPSPKPSDFLLDPYKMT 300

Qy 301 LNPSTKKNKNTGSPDRKPSKSKTDNSLSPLNPKLWCHVHLKKSLSGSPLKVNKSNKSK 360
Db 301 LNPSTKKNKNTGSPDRKPSKSKTDNSLSPLNPKLWCHVHLKKSLSGSPLKVNKSNKSK 360
Qy 361 SPEEHLSEMMKMSPNKLTHTNFHIPKKGPPAKPGKHSKPLKAKGRSKGILNQKSTGN 420
Db 361 SPEEHLSEMMKMSPNKLTHTNFHIPKKGPPAKPGKHSKPLKAKGRSKGILNQKSTGN 420
Qy 421 SKSPKGLKTPKTKWKQMTLLDMAKGTQKMTAPRNSGGTPTRTSSKPKHKLPPAALHLIA 480
Db 421 SKSPKGLKTPKTKWKQMTLLDMAKGTQKMTAPRNSGGTPTRTSSKPKHKLPPAALHLIA 480
Qy 481 YYKENKDREKRSALSVCISKARTLLSSDRARLPEELRSILVOKRYELLEHKKWASMS 540
Db 481 YYKENKDREKRSALSVCISKARTLLSSDRARLPEELRSILVOKRYELLEHKKWASMS 540
Qy 541 EQRKEYLKKKEELKKLKEKAKERREKEMLERLEKQRYEDQELTGKLPAPFLVDTP 600
Db 541 EQRKEYLKKKEELKKLKEKAKERREKEMLERLEKQRYEDQELTGKLPAPFLVDTP 600
Qy 601 GLPNTLFGDVAMVVEFLSCYSGLLLPDAQYPTITAVSLMEALSADGGFLYLNRLVILLQ 660
Db 601 GLPNTLFGDVAMVVEFLSCYSGLLLPDAQYPTITAVSLMEALSADGGFLYLNRLVILLQ 660
Qy 661 TLLQTLQDETAEDYGEGLMKLSEIPLTLHSVSELVRLCLRRSDVQSESGSDTDDNKDS 720
Db 658 -LLQTLQDETAEDYGEGLMKLSEIPLTLHSVSELVRLCLRRSDVQSESGSDTDDNKDS 716
Qy 721 AAPEDNEVDQEFLEKLTSEFFELTSEKLIQTLALCHRLIMTYSDVQDHMETROQMSAEL 780
Db 717 AAPEDNEVDQEFLEKLTSEFFELTSEKLIQTLALCHRLIMTYSDVQDHMETROQMSAEL 776
Qy 781 WKERLAVLKEENDKKRAEKOKRKEMEAKNKENGKVENGLKTDKKBIVKPEPQVDEAE 840
Db 777 WKERLAVLKEENDKKRAEKOKRKEMEAKNKENGKVENGLKTDKKBIVKPEPQVDEAE 836
Qy 841 DMISAVKSRLLATQAKEREIOBERMKVKLERQAEERIRKHAAKAFAOEIGIAKAL 900
Db 837 DMISAVKSRLLALQAKEREIOBERMKVKLERQAEERIRKHAAKAFAOEIGIAKAL 896
Qy 901 VMRTPTIGTDNRNHRNRYWLFSDVEPGLFIEKGWVHDSIDYRNFHHCKDHTVSGDEDYCP 960
Db 897 VMRTPTIGTDNRNHRNRYWLFSDVEPGLFIEKGWVHDSIDYRNFHHCKDHTVSGDEDYCP 956
Qy 961 KKANLGNKASMTQHGTA TEVAVETTPKOGONLWFLCDSQKELDELLNCLHPGIESQ 1020
Db 957 KKANLGNKASMTQHGTA TEVAVETTPKOGONLWFLCDSQKELDELLNCLHPGIESQ 1016
Qy 1021 LKERLEKRYODIHSIHLARKPNLGLKSCDGNQELNFLRSDLLIEVATRLQKGLGVYEE 1080
Db 1017 LKERLEKRYODIHSIHLARKPNLGLKSCDGNQELNFLRSDLLIEVATRLQKGLGVYEE 1076
Qy 1081 TSEFEARVILEKLFKFGECVIALQASVIKFKLQGFMAKQKRKLOSESAKTEEVDEE 1140
Db 1077 TSEFEARVILEKLFKFGECVIALQASVIKFKLQGFMAKQKRKLOSESAKTEEVDEE 1136
Qy 1141 KQWVEEAKVASALEKWKTAIREAQTPSRMHVLLGMDACIKWDMSAENARCKVCPKKGED 1200
Db 1137 KQWVEEAKVASALEKWKTAIREAQTPSRMHVLLGMDACIKWDMSAENARCKVCPKKGED 1196
Qy 1201 DKLLCDECKNAFHLCLRPALEVPDGEWQCPACQATARRNSGRNYTEESASEDSED 1260
Db 1197 DKLLCDECKNAFHLCLRPALEVPDGEWQCPACQATARRNSGRNYTEESASEDSED 1256
Qy 1261 DESDEEEEEEEDYEVAGLRPRKTIIRGKHSVIPPAAKSGRRPGKKPHSTRRSQ 1320
Db 1257 DESDEEEEEEEDYEVAGLRPRKTIIRGKHSVIPPAAKSGRRPGKKPHSTRRSQ 1316
Qy 1321 PKAPPVDDAEVDELVLQTKSSRRQSLQKCEILHKIKVYKRESWPFREPVTDDEADY 1380
Db 1317 PKAPPVDDAEVDELVLQTKSSRRQSLQKCEILHKIKVYKRESWPFREPVTDDEADY 1376
Qy 1381 YDVITHPMDFTQVQNKCSGYSRYSQBFLLTDMKQVFTNAEVYNCRCGSHVLSCMVKTEQCL 1440

Db 1377 YDVIHPMDFTQVQNKSCGSRVSQVEFLTDKQVFTNAEVYNCGRSHVLSQMVKTEOQL 1436
QY 1441 VVLLHKHLPCHPYVVRKRKKFPDRLAEDGSDSEPAVGOSRDEDRRSRE 1489
Db 1437 VALLHKHLPCHPYVVRKRKKFPDRLAEDGSDSEPAVGQSR--GRQKK 1483

RESULT 4
ID ADF69140
AD69140 standard; protein; 1483 AA.
XX ADF69140;
XX
DT 12-FEB-2004 (first entry)
XX Human MP53 protein sequence SEQ ID NO:110.
DE
XX p53 pathway modulating agent; MP53; p53 modulator; cytostatic;
KW Gene therapy; cancer; human.
XX Homo sapiens.
XX
XX WO2003083047-A2.
XX
XX 09-OCT-2003.
XX
XX 28-FEB-2003; 2003WO-US006025.
XX
XX 01-MAR-2002; 2002US-0361196P.
XX
XX (EXEL-) EXELIXIS INC.
XX Belvin M, Francis-Lang H, Friedman L, Plowman GD, Heuer TS, Li D;
PI Funke RP;
XX
XX WPI; 2003-812540/76.
DR N-PSDB; ADF69196.
XX
XX Identifying a candidate p53 pathway modulating agent for treating e.g.,
PT cancer by contacting an assay system comprising a MP53 polypeptide or
PT nucleic acid with a test agent and detecting a test agent-biased
PT activity.
XX
XX Example; SEQ ID NO 110; 406pp; English.
XX
XX The present invention describes a method for identifying a candidate p53
CC pathway modulating agent, which comprises: (a) providing an assay system
CC comprising a MP53 (modulator of p53) polypeptide or nucleic acid or its
CC fragment or derivative; (b) contacting the assay system with a test agent
CC under conditions where the system provides a reference activity except in
CC the presence of the test agent; and (c) detecting a test agent-biased
CC activity, where a difference between the test agent-biased activity and
CC the reference activity identifies the test agent as a candidate p53
CC pathway modulating agent. Also described: (1) modulating the p53 pathway
CC of a cell; (2) modulating the p53 pathway in a mammalian cell; and (3)
CC diagnosing a disease in a patient. MP53 has cytostatic activity, and can
CC be used in gene therapy. The method is useful for identifying a candidate
CC p53 pathway modulating agent for preparing a composition for diagnosing
CC or treating e.g., cancer. The present sequence represents a human MP53
CC protein, which is used in the exemplification of the present invention.
XX
SQ Sequence 1483 AA;

Query Match 95.9%; Score 7656; DB 7; Length 1483;
Best Local Similarity 98.9%; Pred. No. 0;
Matches 1473; Conservative 2; Mismatches 8; Indels 6; Gaps 2;

QY 1 MAPLLGRKPFPLVNPVLPGEFFFTTIPHTQEAFTREEYEARELYSERIWTCKSTGSSQL 60
Db 1 MAPLLGRKPFPLVNPVLPGEFFFTTIPHTQEAFTREEYEARELYSERIWTCKSTGSSQL 60
QY 61 THKEAWEBEQVAELLKEEFPAPWYBKLVLVEMVHNHTASLEKLVDTAWLIMTKYAVGESC 120

Db 61 THKEAWEBEQVAELLKEEFPAPWYBKLVLVEMVHNHTASLEKLVDTAWLIMTKYAVGESC 120
QY 121 DFEVCKEKMLVKVIKIHIPLEKVDDEATEKSGDGCADSPSSDKENSSQIAQDHOKKETT 180
Db 121 DFEVCKEKMLVKVIKIHIPLEKVDDEATEKSGDGCADSPSSDKENSSQIAQDHOKKETT 180
QY 181 KEDEGRRESINDRARRSPRKLPSTSLKGERKWAPPKFLPHKYDVVKLQNEDKIISNPADS 240
Db 181 KEDEGRRESINDRARRSPRKLPSTSLKGERKWAPPKFLPHKYDVVKLQNEDKIISNPADS 240
QY 241 LIRTRPPNKEIVRYFIRHNALRAGTGENAPWVVEDELVKKYSLSKPSKFSDFLLDPKYMT 300
Db 241 LIRTRPPNKEIVRYFIRHNALRAGTGENAPWVVEDELVKKYSLSKPSKFSDFLLDPKYMT 300
QY 301 LNPSTKRKNTGSPDRKPSKSKTDNSSLSSPLNPKLWCHVHLKSLSGSPLVKVNSKNSK 360
Db 301 LNPSTKRKNTGSPDRKPSKSKTDNSSLSSPLNPKLWCHVHLKSLSGSPLVKVNSKNSK 360
QY 361 SPEEHLEEMKMSPNKLTNPHIPKGGPPAKKPGKHSKDKPLKAKGRSKGILNGQKSTGN 420
Db 361 SPEEHLEEMKMSPNKLTNPHIPKGGPPAKKPGKHSKDKPLKAKGRSKGILNGQKSTGN 420
QY 421 SKSPKKGKLTPTKTKMQMTLLDMAKGTQKMTAPRNSGGTPTRTSKPHKHLPPAALHLIA 480
Db 421 SKSPKKGKLTPTKTKMQMTLLDMAKGTQKMTAPRNSGGTPTRTSKPHKHLPPAALHLIA 480
QY 481 YYKENKDRPKSALS CVISKTARLLSSEDRARLPEELRSLVOKRYELLEHKRWASMS 540
Db 481 YYKENKDRPKSALS CVISKTARLLSSEDRARLPEELRSLVOKRYELLEHKRWASMS 540
QY 541 EORKEYLKKREELKKLKERAKERREKEMLERLEKQKRYEYDQELTGKLPAPRLVDTP 600
Db 541 EORKEYLKKREELKKLKERAKERREKEMLERLEKQKRYEYDQELTGKLPAPRLVDTP 600
QY 601 GLPNTLFGDVAMVVEFLSCYSGLLLPDAQYPTTAVSLMEALSADGGFYLNRVLVILQ 660
Db 601 GLPNTLFGDVAMVVEFLSCYSGLLLPDAQYPTTAVSLMEALSADGGFYLNRVLVILQ 660
QY 661 TLLQTLLODEAEDYGEIGMKLSEIPLTLHVSSELVRLCLRSDVDQESSEGGSDTDN 720
Db 658 -LLQTLLODEAEDYGEIGMKLSEIPLTLHVSSELVRLCLRSDVDQESSEGGSDTDN 716
QY 721 AAFEDNEVDQDFLEKLETSEFFELTSEBKQLTALTALCHRLMTYSVDHMETRQOMS 780
Db 717 AAFEDNEVDQDFLEKLETSEFFELTSEBKQLTALTALCHRLMTYSVDHMETRQOMS 776
QY 781 WKERLAVLKEENDKGRABKQKREKEMAKNKNKENGKVENGLGKTDRKKRIVKFPFQV 840
Db 777 WKERLAVLKEENDKGRABKQKREKEMAKNKNKENGKVENGLGKTDRKKRIVKFPFQV 836
QY 841 DMISAVKSRRLLAIOAKKERETQEREMVKLERQAEERIRKHAARAKAFQEGTAKAKL 900
Db 837 DMISAVKSRRLLAIOAKKERETQEREMVKLERQAEERIRKHAARAKAFQEGTAKAKL 896
QY 901 VMRRTPIGTDNRNHRNYLFSDEVPGLFTEKGWVHDSIDYRFNHHCKDHTVSGDE 960
Db 897 VMRRTPIGTDNRNHRNYLFSDEVPGLFTEKGWVHDSIDYRFNHHCKDHTVSGDE 956
QY 961 KXANLGKNASMNTQHGATATEVAVETTPKQGNLWFLCDSQKELDELNCLHPQGIRESQ 1020
Db 957 KXANLGKNASMNTQHGATATEVAVETTPKQGNLWFLCDSQKELDELNCLHPQGIRESQ 1016
QY 1021 LKERLEKEYODIHSIHILARKPNGLKSCDGNOLNPLRSLIEVATRLQGGGYVEE 1080
Db 1017 LKERLEKEYODIHSIHILARKPNGLKSCDGNOLNPLRSLIEVATRLQGGGYVEE 1076
QY 1081 TSEFARVISLEKLDGFCVIALQASVTKFQGFMAPKQKRRKLQSDSDSAKTEVDSE 1140
Db 1077 TSEFARVISLEKLDGFCVIALQASVTKFQGFMAPKQKRRKLQSDSDSAKTEVDSE 1136
QY 1141 KQWEEAKVASALEKWKTAIREAQFTSRMVLGLMLDACIKWDMSAENARCKVCPKKGED 1200

Db 1137 KQVEEAKVASALEKWKTAIREAQTFSRMVLVGLMDACIKWDSNAENACKVCRKKGED 1196
Qy 1201 DKLLCDECNKAFHLCFLRPALYEVDPGEWQCPACQATARNRSGRNYTSESASEDSED 1260
Db 1197 DKLLCDECNKAFHLCFLRPALYEVDPGEWQCPACQATARNRSGRNYTSESASEDSED 1256
Qy 1261 DESDEEEEEEEDYEVAGLRPRKKTIRGKHSVIPPAAARGRPKPKPHSTRSQ 1320
Db 1257 DESDEEEEEEEDYEVAGLRPRKKTIRGKHSVIPPAAARGRPKPKPHSTRSQ 1316
Qy 1321 KPAPVDDAEVDLQTKRSSRQSLQKCEEILHKIVKRYFSPREPVTREADEY 1380
Db 1317 KPAPVDDAEVDLQTKRSSRQSLQKCEEILHKIVKRYFSPREPVTREADEY 1376
Qy 1381 YDVITHPMDFTQVQKSCGYSVQEFITDMKQVFTNAEVNCRGSHVLSQVKTQCL 1440
Db 1377 YDVITHPMDFTQVQKSCGYSVQEFITDMKQVFTNAEVNCRGSHVLSQVKTQCL 1436
Qy 1441 VVLLHKHLPGHYPVRRKPKPDRLAEDGDESEPEAVGQSDRDRSRE 1489
Db 1437 VALLHKHLPGHYPVRRKPKPDRLAEDGDESEPEAVGQSR--GRRQKK 1483

RESULT 5

ID ABR41364 standard; protein; 1052 AA.

XX ABR41364;

AC ABR41364;

XX 02-JUN-2003 (first entry)

XX Human DITHP transcription factor.

XX Human; dithp; diagnostic and therapeutic polynucleotide; diagnosis;
KW cancer; cell proliferative disorder; autoimmune disorder;
KW inflammatory disorder; infection; hormonal disorder; metabolic disorder;
KW neurological disorder; gastrointestinal disorder; transport disorder;
KW connective tissue disorder; drug screening; proteome analysis;
KW gene therapy; antisense therapy; genotyping; transgenic animal; knock in;
KW disease model; toxicological testing; transcript imaging;
KW transcription factor.

XX Homo sapiens.

XX WO200297031-A2.

XX 05-DEC-2002.

XX 27-MAR-2002; 2002WO-US010056.

XX 28-MAR-2001; 2001US-0279619P.

XX 29-MAR-2001; 2001US-0280067P.

XX 16-MAY-2001; 2001US-0291280P.

XX 17-MAY-2001; 2001US-0291829P.

XX 19-JUN-2001; 2001US-0299428P.

XX 20-JUN-2001; 2001US-0299776P.

XX (INCY-) INCYTE GENOMICS INC.

XX Daffo A, Jones AL, Tran AB, Dahl CR, Gietzen D, Chinn J;

PI Dufour GE, Hillman JL, Yu JY, Tuason O, Yap PE, Amshey SR;

PI Daughtery SC, Dam TC, Liu TF, Nguyen DA, Kleefeld Y, Gerstin EH;

PI Pexalta CH, David MH, Lewis SA, Chen AJ, Panzer SR, Harris B;

PI Flores V, Marwaha R, Lo A, Lan RY, Urashka ME;

XX WPI; 2003-129518/12.

DR N-PSDB; ACC46304.

XX Novel human diagnostic and therapeutic polypeptide useful for identifying
PT test compound which specifically binds to a polypeptide encoded by human

PT diagnostic and therapeutic polynucleotide, and to induce antibodies.
XX Claim 27; SEQ ID NO 899; 591pp; English.
CC The invention relates to novel human diagnostic and therapeutic
CC polynucleotides designated dithp (ACC46080-ACC46749) and to their encoded
CC proteins (DITHP; ABR41136-ABR41812). The invention also relates to
CC polynucleotide sequences at least 90% identical to the dithp cDNA
CC sequences of the invention; recombinant vectors, host cells and
CC transgenic organisms comprising a dithp nucleic acid sequence; the
CC recombinant production of DITHP proteins; antibodies specific for DITHP
CC proteins; microarrays comprising dithp nucleic acid sequences; methods of
CC detecting dithp nucleotide and protein sequences; methods of screening
CC for compounds which specifically bind a DITHP protein; and methods of
CC assessing the toxicity of test compounds using a dithp hybridisation
CC probe. Dithp nucleic acid sequences and DITHP proteins may be used in the
CC diagnosis of a wide variety of conditions including cancer and other cell
CC proliferative disorders; autoimmune or inflammatory disorders; bacterial,
CC viral, fungal or parasitic infections; hormonal disorders; metabolic
CC disorders; neurological disorders; gastrointestinal disorders; transport
CC disorders; and connective tissue disorders. They may also be used to
CC screen for modulators of protein activity or gene expression. DITHP
CC proteins can additionally be used in analysis of the proteome of a tissue
CC or cell type and to induce antibodies. The dithp nucleic acids are
CC additionally useful in somatic or germline gene therapy of the disorders
CC mentioned above, as a source of antisense sequences, as a source of
CC probes and primers, in genotyping and identification of individuals, in
CC the generation of transgenic animal models of human disease or knock in
CC humanised animals, in toxicological testing, and in transcript imaging.
CC The present sequence represents a DITHP protein which has transcription
CC factor activity. Note: The sequence data for this patent did not form
CC part of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX Sequence 1052 AA;

Query Match 66.5%; Score 5307; DB 6; Length 1052;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 1034; Conservative 0; Mismatches 1; Indels 4; Gaps 1;
Qy 140 LEKVDEATEKKS DGACDSSPSDKENS SQAQDHQKKE TVVKEDEGRRES INDRARRSPR 199
Db 7 LEKVDEATEKKS DGACDSSPSDKENS SQAQDHQKKE TVVKEDEGRRES INDRARRSPR 66
Qy 200 KLPTSLKKGKRWAPPKFLPHKYDVKLQNE DKIIISNVPADSLIRTERPNKEIVRYFIRH 259
Db 67 KLPTSLKKGKRWAPPKFLPHKYDVKLQNE DKIIISNVPADSLIRTERPNKEIVRYFIRH 126
Qy 260 NALRAGTGENAPWVVEDELVKYSLPSKFSDFLLDPYKMTLNPSTKRKNTGSPDRKPSK 319
Db 127 NALRAGTGENAPWVVEDELVKYSLPSKFSDFLLDPYKMTLNPSTKRKNTGSPDRKPSK 186
Qy 320 KSKTDNSSLSSPLNPKLWCHVHLKKSLSGSPKVKNSKNKSPEEHLEEMKMMSPNKLH 379
Db 187 KSKTDNSSLSSPLNPKLWCHVHLKKSLSGSPKVKNSKNKSPEEHLEEMKMMSPNKLH 246
Qy 380 TNFHIPKKGPPAKPGKHSDKPLKAKGRSKGI LNGQSKSTGNSKSPKGLKTPKTKWKQMT 439
Db 247 TNFHIPKKGPPAKPGKHSDKPLKAKGRSKGI LNGQSKSTGNSKSPKGLKTPKTKWKQMT 306
Qy 440 LLDWAKGTQKMTAPRNSGGTPTSSKPHKHLPAALHLIAYYKENKDREKRSALSCVI 499
Db 307 LLDWAKGTQKMTAPRNSGGTPTSSKPHKHLPAALHLIAYYKENKDREKRSALSCVI 366
Qy 500 SKTARLLSSSDRALRPEELRSLVQRYELLEHHKKRWASMSSEQRKEYLKKCKREELKKLX 559
Db 367 SKTARLLSSSDRALRPEELRSLVQRYELLEHHKKRWASMSSEQRKEYLKKCKREELKKLX 426
Qy 560 EKAKERREKEMLEKQRYEDQELTGKNI LPAFLVDTPPEGLPNTLPGDVAMVVEFLSC 619
Db 427 EKAKERREKEMLEKQRYEDQELTGKNI LPAFLVDTPPEGLPNTLPGDVAMVVEFLSC 486
Qy 620 YSGLLLPDAQYPI TAVSLMEALSADKGFVLYNRLVILLQTLQLQDDEIAEDYGEIG 679

Db 61 NSKSPPEHLEEMKMSPNKLTHTNPHI PKKGPAPKPKGKSDKPLKAKGRSGKILNGQKS 120
QY 418 TGNKSPKPKGLTPTKTKQMTLLDMAKGTQKMTAPRNSGGTPTRTSKPKHKLPPAALH 477
Db 121 TGNKSPKPKGLTPTKTKQMTLLDMAKGTQKMTAPRNSGGTPTRTSKPKHKLPPAALH 180
QY 478 LIAYYKENKDRKDSALSCVISTARLLSSDRARLPPELSLVOKRYELLEHKRWAS 537
Db 181 LIAYYKENKDRKDSALSCVISTARLLSSDRARLPPELSLVOKRYELLEHKRWAS 240
QY 538 MSEEORKEYLKKREELKKLKEKAKERREKEMLEKOKRYEDQELTGKNLPAPFLVD 597
Db 241 MSEEORKEYLKKREELKKLKEKAKERREKEMLEKOKRYEDQELTGKNLPAPFLVD 300
QY 598 TPEGLPNTLFGDVAMVVEFLSCYSGLLLPDAQYPITAVSLMEALSADKGGFLYLNRLVI 657
Db 301 TPEGLPNTLFGDVAMVVEFLSCYSGLLLPDAQYPITAVSLMEALSADKGGFLYLNRLVI 360
QY 658 LLQTLLOLQDEIAEDYGELGMLKSEIPLTLHVSSELVRLCLRRSDVQEESEGSDTDN 717
Db 361 ----LLQTLLOLQDEIAEDYGELGMLKSEIPLTLHVSSELVRLCLRRSDVQEESEGSDTDN 416
QY 718 KDSAAFEDNEVQDEFLEKLTSEFFELTSEELKQILTLCHRLMTYSVQDHMETROQMS 777
Db 417 KDSAAFEDNEVQDEFLEKLTSEFFELTSEELKQILTLCHRLMTYSVQDHMETROQMS 476
QY 778 AELWKERLAVLKEENDKKRAEKQKKEAKNKENGKVENGLGKTDRKKRIVKFPQVDT 837
Db 477 AELWKERLAVLKEENDKKRAEKQKKEAKNKENGKVENGLGKTDRKKRIVKFPQVDT 536
QY 838 EADMISAVKSRRLLAIOAKKERETQEREMKVYL 871
Db 537 EADMISAVKSRRLLAIOAKKERETQEREMKGI 570

RESULT 9

ABB28622
ID ABB28622 standard; peptide; 572 AA.
XX AC ABB28622;
XX DT 01-FEB-2002 (first entry)
XX DE Peptide #1273 encoded by breast cell single exon nucleic acid probe.
XX KW Human; microarray; single exon probe; gene expression; breast; disease;
XX KW cancer.
XX OS Homo sapiens.
XX PN WO200157271-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US0000662.
XX PR 04-FEB-2000; 2000US-0180312P.
XX PR 26-MAY-2000; 2000US-0207456P.
XX PR 30-JUN-2000; 2000US-00608408.
XX PR 03-AUG-2000; 2000US-00632366.
XX PR 21-SEP-2000; 2000US-0234687P.
XX PR 27-SEP-2000; 2000US-0236359P.
XX PR 04-OCT-2000; 2000GB-00024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX PA Penn SG, Hanzel DK, Chen W, Rank DR;
XX PI WPI; 2001-496933/54.
XX DR
XX PT New spatially-addressable set of single exon nucleic acid probes, useful
PT for measuring gene expression in sample derived from human breast,
PT comprises number of single exon nucleic acid probes.

XX Claim 27; SEQ ID NO 11590; 327pp + Sequence Listing; English.

XX The invention relates to a spatially-addressable set of single exon
CC nucleic acid probes for measuring gene expression in a sample derived
CC from human breast and B1 4/4 cells. The method involves contacting the
CC probes with a collection of detectably labelled nucleic acids derived
CC from mRNA of human breast, and then measuring the label bound to each
CC probe of the microarray. The probes are useful for verifying the
CC expression of regions of genomic DNA predicted to encode proteins. They
CC are useful for gene discovery, and for determining predisposition and/or
CC prognosing breast disease. Gene expression analysis is useful for
CC assessing the toxicity of chemical agents on cells. The microarray of
CC this invention presents a far greater diversity of probes for measuring
CC gene expression, with far less bias than expressed sequence tag
CC microarrays. The method is suitable for rapid production of functional
CC information from genomic sequence. The present sequence is a peptide
CC encoded by a single exon nucleic acid probe of the invention. Note: The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences

SQ Sequence 572 AA;

Query Match 36.0%; Score 2877; DB 4; Length 572;
Best Local Similarity 98.8%; Pred. No. 5.6e-190;
Matches 567; Conservative 1; Mismatches 2; Indels 4; Gaps 1;

QY 298 YMTLNPSTKRNKTSPPDRKPSKSKTDSNSSLSSPLNPKLWCHVHLKSLSGSPLKVQNSK 357
Db 1 YMTLNPSTKRNKTSPPDRKPSKSKTDSNSSLSSPLNPKLWCHVHLKSLSGSPLKVQNSK 60
QY 358 NSKSPPEHLEEMKMSPNKLTHTNPHI PKKGPAPKPKGKSDKPLKAKGRSGKILNGQKS 417
Db 61 NSKSPPEHLEEMKMSPNKLTHTNPHI PKKGPAPKPKGKSDKPLKAKGRSGKILNGQKS 120
QY 418 TGNKSPKPKGLTPTKTKQMTLLDMAKGTQKMTAPRNSGGTPTRTSKPKHKLPPAALH 477
Db 121 TGNKSPKPKGLTPTKTKQMTLLDMAKGTQKMTAPRNSGGTPTRTSKPKHKLPPAALH 180
QY 478 LIAYYKENKDRKDSALSCVISTARLLSSDRARLPPELSLVOKRYELLEHKRWAS 537
Db 181 LIAYYKENKDRKDSALSCVISTARLLSSDRARLPPELSLVOKRYELLEHKRWAS 240
QY 538 MSEEORKEYLKKREELKKLKEKAKERREKEMLEKOKRYEDQELTGKNLPAPFLVD 597
Db 241 MSEEORKEYLKKREELKKLKEKAKERREKEMLEKOKRYEDQELTGKNLPAPFLVD 300
QY 598 TPEGLPNTLFGDVAMVVEFLSCYSGLLLPDAQYPITAVSLMEALSADKGGFLYLNRLVI 657
Db 301 TPEGLPNTLFGDVAMVVEFLSCYSGLLLPDAQYPITAVSLMEALSADKGGFLYLNRLVI 360
QY 658 LLQTLLOLQDEIAEDYGELGMLKSEIPLTLHVSSELVRLCLRRSDVQEESEGSDTDN 717
Db 361 ----LLQTLLOLQDEIAEDYGELGMLKSEIPLTLHVSSELVRLCLRRSDVQEESEGSDTDN 416
QY 718 KDSAAFEDNEVQDEFLEKLTSEFFELTSEELKQILTLCHRLMTYSVQDHMETROQMS 777
Db 417 KDSAAFEDNEVQDEFLEKLTSEFFELTSEELKQILTLCHRLMTYSVQDHMETROQMS 476
QY 778 AELWKERLAVLKEENDKKRAEKQKKEAKNKENGKVENGLGKTDRKKRIVKFPQVDT 837
Db 477 AELWKERLAVLKEENDKKRAEKQKKEAKNKENGKVENGLGKTDRKKRIVKFPQVDT 536
QY 838 EADMISAVKSRRLLAIOAKKERETQEREMKVYL 871
Db 537 EADMISAVKSRRLLAIOAKKERETQEREMKGI 570

RESULT 10

ABB19248
ID ABB19248 standard; protein; 572 AA.
XX

Query Match 36.0%; Score 2877; DB 4; Length 572;
Best Local Similarity 98.8%; Pred. No. 5.6e-190;
Matches 567; Conservative 1; Mismatches 2; Indels 4; Gaps 1;

QY 298 YMTLNSTKRNKGTSPDRKPSKSKTDNSLSPLNPKLWCHVHLKXSLGSPKLVKNSK 357
DB 1 YMTLNSTKRNKGTSPDRKPSKSKTDNSLSPLNPKLWCHVHLKXSLGSPKLVKNSK 60

QY 358 NSKSPPEHLEEMKMWSPNKLHNFHPPKGGPAKKPGKHSDKPLKAKGRSGILNGQKS 417
DB 61 NSKSPPEHLEEMKMWSPNKLHNFHPPKGGPAKKPGKHSDKPLKAKGRSGILNGQKS 120

QY 418 TGNSSPKKGLTKPTKTKQMTLLDMAKGTQKTRAPRNSGGTPTRTSSPKHKLPPAALH 477
DB 121 TGNSSPKKGLTKPTKTKQMTLLDMAKGTQKTRAPRNSGGTPTRTSSPKHKLPPAALH 180

QY 478 LIAYYKENDREDKRSALSCVISTARLLSSDRARLPBELSLVQRYELLEHKRWAS 537
DB 181 LIAYYKENDREDKRSALSCVISTARLLSSDRARLPBELSLVQRYELLEHKRWAS 240

QY 538 MSEEQRKEYLKKREELKKLKEKAKERREKEMLEKQRYEDQELTGKNLPAPFLVD 597
DB 241 MSEEQRKEYLKKREELKKLKEKAKERREKEMLEKQRYEDQELTGKNLPAPFLVD 300

QY 598 TPEGLPNTLFGDVAMVVEFLSCYSGILLPDAQYPTAVSLMEALSADKGGFLYLNRVLVI 657
DB 301 TPEGLPNTLFGDVAMVVEFLSCYSGILLPDAQYPTAVSLMEALSADKGGFLYLNRVLVI 360

QY 658 LIQTLLOTLQDEIAEDYGELGKMLSEIPLTLHVSSELVRLCLRRSDVOEESGSDTDN 717
DB 361 ----LIQTLLOTLQDEIAEDYGELGKMLSEIPLTLHVSSELVRLCLRRSDVOEESGSDTDN 416

QY 718 KDSAAFEDNEVDQEFLEKLETSEFFELTSEEKLIQILTALCHRLMTYSVQDHMETRQMS 777
DB 417 KDSAAFEDNEVDQEFLEKLETSEFFELTSEEKLIQILTALCHRLMTYSVQDHMETRQMS 476

QY 778 AELWKERLAVLKEENDKGAERKQKREMEAKNKENGKVENGLGKTDKRGRIVKFEPQVDT 837
DB 477 AELWKERLAVLKEENDKGAERKQKREMEAKNKENGKVENGLGKTDKRGRIVKFEPQVDT 536

QY 838 EAEDMISAVKSRRLAIQAKKEREIOEREMKVKL 871
DB 537 EAEDMISAVKSRRLAIQAKKEREIOEREMKGKI 570

RESULT 12
AAM54571
ID AAM54571 standard; protein; 572 AA.
AC AAM54571;
XX
XX
XX
DT 05-NOV-2001 (first entry)
XX
DE Human brain expressed single exon probe encoded protein SEQ ID NO: 26676.
XX
KW Human; brain expressed exon; gene expression analysis; probe; microarray;
KW Alzheimer's disease; multiple sclerosis; schizophrenia; epilepsy; cancer.
XX
OS Homo sapiens.
XX
PN WO200157275-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US000667.
XX
PR 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-0060840B.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.
PA Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-483446/52.
XX Single exon nucleic acid probes for analyzing gene expression in human
PT brains.
PT
XX
XX Example 4; SEQ ID NO 26676; 650pp + Sequence Listing; English.
XX
CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC brain. They can be used to measure gene expression in brain cell samples,
CC which may enable the diagnosis and improved treatment of nervous system
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC epilepsy and cancers. The present sequence is a protein encoded by one of
CC the probes of the invention
XX
SQ Sequence 572 AA;
Query Match 36.0%; Score 2877; DB 4; Length 572;
Best Local Similarity 98.8%; Pred. No. 5.6e-190;
Matches 567; Conservative 1; Mismatches 2; Indels 4; Gaps 1;

QY 298 YMTLNSTKRNKGTSPDRKPSKSKTDNSLSPLNPKLWCHVHLKXSLGSPKLVKNSK 357
DB 1 YMTLNSTKRNKGTSPDRKPSKSKTDNSLSPLNPKLWCHVHLKXSLGSPKLVKNSK 60

QY 358 NSKSPPEHLEEMKMWSPNKLHNFHPPKGGPAKKPGKHSDKPLKAKGRSGILNGQKS 417
DB 61 NSKSPPEHLEEMKMWSPNKLHNFHPPKGGPAKKPGKHSDKPLKAKGRSGILNGQKS 120

QY 418 TGNSSPKKGLTKPTKTKQMTLLDMAKGTQKTRAPRNSGGTPTRTSSPKHKLPPAALH 477
DB 121 TGNSSPKKGLTKPTKTKQMTLLDMAKGTQKTRAPRNSGGTPTRTSSPKHKLPPAALH 180

QY 478 LIAYYKENDREDKRSALSCVISTARLLSSDRARLPBELSLVQRYELLEHKRWAS 537
DB 181 LIAYYKENDREDKRSALSCVISTARLLSSDRARLPBELSLVQRYELLEHKRWAS 240

QY 538 MSEEQRKEYLKKREELKKLKEKAKERREKEMLEKQRYEDQELTGKNLPAPFLVD 597
DB 241 MSEEQRKEYLKKREELKKLKEKAKERREKEMLEKQRYEDQELTGKNLPAPFLVD 300

QY 598 TPEGLPNTLFGDVAMVVEFLSCYSGILLPDAQYPTAVSLMEALSADKGGFLYLNRVLVI 657
DB 301 TPEGLPNTLFGDVAMVVEFLSCYSGILLPDAQYPTAVSLMEALSADKGGFLYLNRVLVI 360

QY 658 LIQTLLOTLQDEIAEDYGELGKMLSEIPLTLHVSSELVRLCLRRSDVOEESGSDTDN 717
DB 361 ----LIQTLLOTLQDEIAEDYGELGKMLSEIPLTLHVSSELVRLCLRRSDVOEESGSDTDN 416

QY 718 KDSAAFEDNEVDQEFLEKLETSEFFELTSEEKLIQILTALCHRLMTYSVQDHMETRQMS 777
DB 417 KDSAAFEDNEVDQEFLEKLETSEFFELTSEEKLIQILTALCHRLMTYSVQDHMETRQMS 476

QY 778 AELWKERLAVLKEENDKGAERKQKREMEAKNKENGKVENGLGKTDKRGRIVKFEPQVDT 837
DB 477 AELWKERLAVLKEENDKGAERKQKREMEAKNKENGKVENGLGKTDKRGRIVKFEPQVDT 536

QY 838 EAEDMISAVKSRRLAIQAKKEREIOEREMKVKL 871
DB 537 EAEDMISAVKSRRLAIQAKKEREIOEREMKGKI 570

RESULT 13
ABG48640
ID ABG48640 standard; peptide; 572 AA.
XX
XX ABG48640;
XX

```

DT 25-FEB-2003 (first entry)
XX Human liver peptide, SEQ ID NO 27288.
DE Human; liver; cirrhosis; hyperlipoproteinaemia; hyperlipidaemia;
XX hypercholesterolaemia; coronary heart disease.
KW Homo sapiens.
OS Homo sapiens.
XX WO2000157273-A2.
PN 09-AUG-2001.
PD 30-JAN-2001; 2001WO-US000664.
PF 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
PA Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488898/53.
XX Human genome-derived single exon nucleic acid probes useful for analyzing
XX gene expression in human adult liver.
XX Claim 27; SEQ ID NO 27288; 658pp; English.
XX The invention relates to a single exon nucleic acid probe (SENP) (I) for
XX measuring human gene expression in a sample derived from human adult
XX liver, comprising one of 13109 defined nucleotide sequences given in the
XX specification (or complements/ fragments). The probe hybridises at high
XX stringency to a nucleic acid molecule expressed in the human adult liver.
XX (I) may be used for predicting, measuring and displaying gene expression
XX in samples derived from human adult liver. The genes identified may be
XX involved in genetic liver diseases such as cirrhosis,
XX hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which is
XX associated with coronary heart disease. ABG47348-ABG59930 represent human
XX liver single exon encoded peptides of the invention. Note: the sequence
XX information for this patent does not appear in the printed specification
XX but was obtained in electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX Sequence 572 AA;
SQ
Query Match 36.0%; Score 2877; DB 4; Length 572;
Best Local Similarity 98.8%; Pred. No. 5.6e-190;
Matches 567; Conservative 1; Mismatches 2; Indels 4; Gaps 1;
QY 298 YMTLNPSTKRNTGSPDRKPKSKSTDNSSLSPLNPKLWCHVHLKSLSGSPLKVNKSK 357
DB 1 YMTLNPSTKRNTGSPDRKPKSKSTDNSSLSPLNPKLWCHVHLKSLSGSPLKVNKSK 60
QY 358 NSKSPPEEHLKEMKMSNKLHTNPHIPKGGPPAKKPGKSDKPLKAKGRSKGIINGOKS 417
DB 61 NSKSPPEEHLKEMKMSNKLHTNPHIPKGGPPAKKPGKSDKPLKAKGRSKGIINGOKS 120
QY 418 TGNKSPKKGKLTPTKTKMQMTLLDMAKGTQMTAPRNSGGTPTRTSKPHKHLPPAALH 477
DB 121 TGNKSPKKGKLTPTKTKMQMTLLDMAKGTQMTAPRNSGGTPTRTSKPHKHLPPAALH 180
QY 478 LIAYYKENKDREKRSALSCVISTARLLSSEDRARLPEELRSLVQKYEYELLEHKRWAS 537
DB 181 LIAYYKENKDREKRSALSCVISTARLLSSEDRARLPEELRSLVQKYEYELLEHKRWAS 240
QY 538 MSEQRKEYLKKREELKKLKEKAKERREKEMLERLEKQKRYEDQELTGKNLPAFRLVD 597

```

```

DB 241 MSEQRKEYLKKREELKKLKEKAKERREKEMLERLEKQKRYEDQELTGKNLPAFRLVD 300
QY 598 TPEGLPNTLFGDVAMVVEFLSCYSGLLLPDAQYPTITAVSLMEALSADKGGFLYLNRLVI 657
DB 301 TPEGLPNTLFGDVAMVVEFLSCYSGLLLPDAQYPTITAVSLMEALSADKGGFLYLNRLVI 360
QY 658 LLQTLLOTLQDETAEDYVGEGLGMKLSIPLTLHVSSELVRLCLRRSDVQEESESGSDTDN 717
DB 361 ----LLQTLLOTLQDETAEDYVGEGLGMKLSIPLTLHVSSELVRLCLRRSDVQEESESGSDTDN 416
QY 718 KDSAAFEDNEVQDFLEKLETSEFFELTSEKQLIQTALCHRLMTYSVQDHMETRQOMS 777
DB 417 KDSAAFEDNEVQDFLEKLETSEFFELTSEKQLIQTALCHRLMTYSVQDHMETRQOMS 476
QY 778 AELWKERLAVLKEENDKKRAEKQKRAEKMEAKNKGKVENGLGKTDRKRIVKPEPQVDT 837
DB 477 AELWKERLAVLKEENDKKRAEKQKRAEKMEAKNKGKVENGLGKTDRKRIVKPEPQVDT 536
QY 838 EADMISAVKSRRLLAIOAKKEREIQEREMKVKL 871
DB 537 EADMISAVKSRRLLAIOAKKEREIQEREMKVKI 570
RESULT 14
AAM02563
ID AAM02563 standard; protein; 572 AA.
XX AC AAM02563;
XX 09-OCT-2001 (first entry)
XX Peptide #1245 encoded by probe for measuring breast gene expression.
XX Probe; human; breast disease; breast cancer; development disorder;
XX inflammatory disease; proliferative breast disease; non-carcinoma tumour.
XX Homo sapiens.
XX WO2000157270-A2.
XX 09-AUG-2001.
XX 29-JAN-2001; 2001WO-US000661.
XX 04-FEB-2000; 2000US-0180312P.
XX 26-MAY-2000; 2000US-0207456P.
XX 30-JUN-2000; 2000US-00608408.
XX 03-AUG-2000; 2000US-00632366.
XX 21-SEP-2000; 2000US-0234687P.
XX 27-SEP-2000; 2000US-0236359P.
XX 04-OCT-2000; 2000GB-00024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
PA Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-476286/51.
XX Novel single exon nucleic acid probe used to measuring gene expression in
XX a human breast.
XX Claim 27; SEQ ID NO 11303; 322pp; English.
XX The present invention relates to novel single exon nucleic acid probes
XX (see AAI00010-AAI10067). The present sequence is a peptide encoded by one
XX such probe. The probes are useful for measuring human gene expression in
XX a human breast sample, where the probe hybridises at high stringency to a
XX nucleic acid expressed in the human breast. The probes are useful for
XX predicting, diagnosing, grading, staging, monitoring and prognosing
XX diseases of the human breast, particularly those diseases with polygenic
XX aetiology. The diseases include: breast cancer, disorders of development,
XX inflammatory diseases of the breast, fibrocystic changes, proliferative
XX breast disease and non-carcinoma tumours. Note: The sequence data for

```

CC this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 572 AA;

Query Match 36.0%; Score 2877; DB 4; Length 572;
Best Local Similarity 98.8%; Pred. No. 5.6e-190;
Matches 567; Conservative 1; Mismatches 2; Indels 4; Gaps 1;

QY 298 YMTLNPSTKRKNTGSPDRKPSKSTDNSSLSPLNPKLWCHVHLKSLSGSPLKYNKSK 357
DB 1 YMTLNPSTKRKNTGSPDRKPSKSTDNSSLSPLNPKLWCHVHLKSLSGSPLKYNKSK 60

QY 358 NSKSPPEHLEEMKQWMSPNKLTNPHI PKKGPAPKPKGHSKDKPLKAKGRSGIINGOKS 417
DB 61 NSKSPPEHLEEMKQWMSPNKLTNPHI PKKGPAPKPKGHSKDKPLKAKGRSGIINGOKS 120

QY 418 TGNKSPKPKGLTKPTKMKQMTLLDVAKGTOQMTAPRNSGCTPTTSKPKHKLPPAALH 477
DB 121 TGNKSPKPKGLTKPTKMKQMTLLDVAKGTOQMTAPRNSGCTPTTSKPKHKLPPAALH 180

QY 478 LIAYYKENKORDKESALSCVISTARTLLSSDRARLPPEELSLVQKRYELLEHKRWAS 537
DB 181 LIAYYKENKORDKESALSCVISTARTLLSSDRARLPPEELSLVQKRYELLEHKRWAS 240

QY 538 MSEEORKEVLYKKREELKKLKEKAKERREKEMLERLEKQKRYEDQELTGKLPAPFLVD 597
DB 241 MSEEORKEVLYKKREELKKLKEKAKERREKEMLERLEKQKRYEDQELTGKLPAPFLVD 300

QY 598 TPEGLPNTLFGDVAMVVEFLSCYSGLLLPDAQYPTAVSLMEALSADKGGFLYLNRLVI 657
DB 301 TPEGLPNTLFGDVAMVVEFLSCYSGLLLPDAQYPTAVSLMEALSADKGGFLYLNRLVI 360

QY 658 LLQTLQTLQDEIAEDYGEIEMKLSLSEPLTHSVSELVRLCLRRSDVOERSEGSDDTN 717
DB 361 ---LLQTLQTLQDEIAEDYGEIEMKLSLSEPLTHSVSELVRLCLRRSDVOERSEGSDDTN 416

QY 718 KDSAAFENEVDQEFLEKLETSEFFELTSEELQLITLALCHRLMTYSVDHMETRQOMS 777
DB 417 KDSAAFENEVDQEFLEKLETSEFFELTSEELQLITLALCHRLMTYSVDHMETRQOMS 476

QY 778 AELWKEKRLAVLKEENDKRAEKQKREMEAKNKENGKVENGLGKTDKRRIVKPEPQVDT 837
DB 477 AELWKEKRLAVLKEENDKRAEKQKREMEAKNKENGKVENGLGKTDKRRIVKPEPQVDT 536

QY 838 EAEDMISAVKSRRLIAIAQKKEREIQEREMKVKL 871
DB 537 EAEDMISAVKSRRLIAIAQKKEREIQEREMKVKI 570

RESULT 15
ID ABG36635 standard; peptide; 572 AA.
AC ABG36635;
XX
XX
DT 19-AUG-2002 (first entry)
DE
DE Human peptide encoded by genome-derived single exon probe SEQ ID 26300.
KW Human; single exon probe; asthma; lung cancer; COPD; ILD;
KW chronic obstructive pulmonary disease; interstitial lung disease;
KW familial idiopathic pulmonary fibrosis; neurofibromatosis;
KW tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;
KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;
KW pulmonary histiocytosis; lymphangioleiomyomatosis; Karagenar syndrome;
KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
KW primary ciliary dyskinesia; pulmonary hypertension;
KW hyaline membrane disease.
OS Homo sapiens.
XX

PN WO200186003-A2.
XX 15-NOV-2001.
XX
XX 30-JAN-2001; 2001WO-US000665.
XX
XX 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2002-114183/15.
XX
XX Spatially-addressable set of single exon nucleic acid probes, used to
XX measure gene expression in human lung samples.
XX
XX Claim 27; SEQ ID NO 26300; 634pp; English.
XX
XX The invention relates to a spatially-addressable set of single exon
XX nucleic acid probes for measuring gene expression in a sample derived
XX from human lung comprising single exon nucleic acid probes having one of
XX 12614 nucleic acid sequences mentioned in the specification, or their
XX complements or the 12387 open reading frames derived from the 12614
XX probes. Also included are a microarray comprising the novel set of probes
XX; the novel set of probes which hybridise at high stringency to a nucleic
XX acid expressed in the human lung; measuring gene expression in a sample
XX derived from human lung, comprising (a) contacting the array with a
XX collection of detectably labeled nucleic acids derived from human lung
XX mRNA, and (b) measuring the label detectably bound to each probe of the
XX array; identifying exons in a eukaryotic genome, comprising (a)
XX algorithmically predicting at least one exon from genomic sequences of
XX the eukaryote; and (b) detecting specific hybridisation of detectably
XX labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,
XX having a fragment identical to the predicted exon, the probe is included
XX in the above mentioned microarray; assigning exons to a single gene,
XX comprising (a) identifying exons from genomic sequence by the method
XX above and (b) measuring the expression of each of the exons in several
XX tissues and/or cell types using hybridisation to a single exon
XX microarrays having a probe with the exon, where a common pattern of
XX expression of the exons in the tissues and/or cell types indicates that
XX the exons should be assigned to a single gene; a peptide comprising one
XX of 12011 sequences, mentioned in the specification, or encoded by the
XX probes/open reading frames (ORF). The probes are used for gene expression
XX analysis, and for identifying exons in a gene, particularly using human
XX lung derived mRNA and for the study of lung diseases such as asthma, lung
XX cancer, chronic obstructive pulmonary disease (COPD), interstitial lung
XX disease (ILD), familial idiopathic pulmonary fibrosis, neurofibromatosis,
XX tuberous sclerosis, Gaucher's disease, Niemann-Pick disease, Hermansky-
XX Pudlak syndrome, sarcoidosis, pulmonary haemosiderosis, pulmonary
XX histiocytosis, lymphangioleiomyomatosis, pulmonary alveolar proteinosis,
XX Karagenar syndrome, fibrocystic pulmonary dysplasia, primary ciliary
XX dyskinesia, pulmonary hypertension and hyaline membrane disease. The
XX present sequence is a peptide/protein encoded by a single exon probe of
XX the invention. Note: The sequence data for this patent did not form part
XX of the printed specification, but was obtained in electronic format
XX directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 572 AA;

Query Match 36.0%; Score 2877; DB 5; Length 572;
Best Local Similarity 98.8%; Pred. No. 5.6e-190;
Matches 567; Conservative 1; Mismatches 2; Indels 4; Gaps 1;

QY 298 YMTLNPSTKRKNTGSPDRKPSKSTDNSSLSPLNPKLWCHVHLKSLSGSPLKYNKSK 357
DB 1 YMTLNPSTKRKNTGSPDRKPSKSTDNSSLSPLNPKLWCHVHLKSLSGSPLKYNKSK 60

Qy 358 NSKSPPEHLEEMKMSPNKLTHTNPHIPKGGPPAKPKGKSHDKPLKAKGRSKGILNGOKS 417
Db 61 NSKSPPEHLEEMKMSPNKLTHTNPHIPKGGPPAKPKGKSHDKPLKAKGRSKGILNGOKS 120
Qy 418 TGNKSPKKGKLTPTKTKMQMTLLDMAKGTQKTRAPNSGGTPTRTSSKPHKHLPPAALH 477
Db 121 TGNKSPKKGKLTPTKTKMQMTLLDMAKGTQKTRAPNSGGTPTRTSSKPHKHLPPAALH 180
Qy 478 LIAYYKENKDRKRSALSCVISTARTLLSSEDRARLPPEELRSVQKRYELLEHKKRWAS 537
Db 181 LIAYYKENKDRKRSALSCVISTARTLLSSEDRARLPPEELRSVQKRYELLEHKKRWAS 240
Qy 538 MSEQRKRYLKKREELKKLKEKAKERREKEMLERLEKQRYEDQELTGKNLPAFRLLVD 597
Db 241 MSEQRKRYLKKREELKKLKEKAKERREKEMLERLEKQRYEDQELTGKNLPAFRLLVD 300
Qy 598 TPEGLPNTLFGDVAMVVEFLSCYSGLLLPDAQYPTAVSLMEALSADKGGFLYLNRVLVI 657
Db 301 TPEGLPNTLFGDVAMVVEFLSCYSGLLLPDAQYPTAVSLMEALSADKGGFLYLNRVLVI 360
Qy 658 LLQTLQTLQDETAEDYGELGMKLSSEIPLTLHVSSELVRLCLRRSDVQESSESGDITDND 717
Db 361 ---LLQTLQDETAEDYGELGMKLSSEIPLTLHVSSELVRLCLRRSDVQESSESGDITDND 416
Qy 718 KDSAAFEDNEVDREFLEKLTSEPFELTSEBKQILTLALCHRIILMTYSVQDHMETROOMS 777
Db 417 KDSAAFEDNEVDREFLEKLTSEPFELTSEBKQILTLALCHRIILMTYSVQDHMETROOMS 476
Qy 778 AELWKERLAVLKEENDKKRAEKQKREKEMAKNKGKVENGLGKTDRKKRVKPEPQVDT 837
Db 477 AELWKERLAVLKEENDKKRAEKQKREKEMAKNKGKVENGLGKTDRKKRVKPEPQVDT 536
Qy 838 EAEDMISAVKRRLLIAIOAKKERIQREREMKVL 871
Db 537 EAEDMISAVKRRLLIAIOAKKERIQREREMKGI 570
RESULT 16
AAM18320
ID AAM18320 standard; protein; 560 AA.
XX
AC AAM18320;
XT 12-OCT-2001 (first entry)
DE Peptide #4754 encoded by probe for measuring cervical gene expression.
KW Probe; human; microarray; gene expression; cervical epithelial cell;
KW cervical cancer.
XX Homo sapiens.
XX WO200157278-A2.
XX 09-AUG-2001.
XX 30-JAN-2001; 2001WO-US000670.
XX 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
FI WPI; 2001-488901/53.
XX

PT Human genome-derived single exon nucleic acid probes useful for analyzing
PT gene expression in human cervical epithelial cells.
XX Claim 27; SEQ ID NO 23146; 487pp; English.
CC The present invention relates to human single exon nucleic acid probes
CC (SNP: see AAI10068-AA128459). The present sequence is a peptide encoded
CC by one such probe. The SNPs are derived from human HeLa cells. The SNPs
CC can be used to produce a single exon microarray, which can be used for
CC measuring human gene expression in a sample derived from human cervical
CC epithelial cells. By measuring gene expression, the probes are therefore
CC useful in grading and/or staging of diseases of the cervix, notably
CC cervical cancer. Note: The sequence data for this patent did not form
CC part of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX Sequence 560 AA;
SQ Query Match 35.5%; Score 2835; DB 4; Length 560;
Best Local Similarity 99.1%; Pred. No. 4.3e-187;
Matches 559; Conservative 0; Mismatches 1; Indels 4; Gaps 1;
Qy 305 TKRNTGSPDRPKSPKSKTNSLSPLNPKLWCHVHLKKSLSGSLPKVKNKSKSPPEE 364
Db 1 TKRNTGSPDRPKSPKSKTNSLSPLNPKLWCHVHLKKSLSGSLPKVKNKSKSPPEE 60
Qy 365 HLEEMKMSPNKLTHTNPHIPKGGPPAKPKGKSHDKPLKAKGRSKGILNGOKS 424
Db 61 HLEEMKMSPNKLTHTNPHIPKGGPPAKPKGKSHDKPLKAKGRSKGILNGOKS 120
Qy 425 KKGKLTPTKTKMQMTLLDMAKGTQKTRAPNSGGTPTRTSSKPHKHLPPAALH 484
Db 121 KKGKLTPTKTKMQMTLLDMAKGTQKTRAPNSGGTPTRTSSKPHKHLPPAALH 180
Qy 485 NKDREKRSALSCVISTARTLLSSEDRARLPPEELRSVQKRYELLEHKKRWAS 544
Db 181 NKDREKRSALSCVISTARTLLSSEDRARLPPEELRSVQKRYELLEHKKRWAS 240
Qy 545 EYLKKEEELKKLKEKAKERREKEMLERLEKQRYEDQELTGKNLPAFRLLVD 604
Db 241 EYLKKEEELKKLKEKAKERREKEMLERLEKQRYEDQELTGKNLPAFRLLVD 300
Qy 605 TLFGDVAMVVEFLSCYSGLLLPDAQYPTAVSLMEALSADKGGFLYLNRVLVI 664
Db 301 TLFGDVAMVVEFLSCYSGLLLPDAQYPTAVSLMEALSADKGGFLYLNRVLVI 356
Qy 665 TLLQDEIAEDYGELGMKLSSEIPLTLHVSSELVRLCLRRSDVQESSESGDITDND 724
Db 357 TLLQDEIAEDYGELGMKLSSEIPLTLHVSSELVRLCLRRSDVQESSESGDITDND 416
Qy 725 DNEVDREFLEKLTSEPFELTSEBKQILTLALCHRIILMTYSVQDHMETROOMS 784
Db 417 DNEVDREFLEKLTSEPFELTSEBKQILTLALCHRIILMTYSVQDHMETROOMS 476
Qy 785 LAVLKEENDKKRAEKQKREKEMAKNKGKVENGLGKTDRKKRVKPEPQVDT 844
Db 477 LAVLKEENDKKRAEKQKREKEMAKNKGKVENGLGKTDRKKRVKPEPQVDT 536
Qy 845 AVKSRRLIAIOAKKERIQREREMK 868
Db 537 AVKSRRLIAIOAKKERIQREREMK 560
RESULT 17
ABB37354
ID ABB37354 standard; peptide; 560 AA.
XX
AC ABB37354;
XX
DT 04-FEB-2002 (first entry)
XX
DE Peptide #4860 encoded by human foetal liver single exon probe.
XX

KW Human; foetal liver; gene expression; single exon nucleic acid probe.
XX Homo sapiens.
XX WO200157277-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US000669.
XX PR 04-FEB-2000; 2000US-0180312P.
XX PR 26-MAY-2000; 2000US-0207456P.
XX PR 30-JUN-2000; 2000US-00608408.
XX PR 03-AUG-2000; 2000US-00632366.
XX PR 21-SEP-2000; 2000US-0234687P.
XX PR 27-SEP-2000; 2000US-0236359P.
XX PR 04-OCT-2000; 2000GB-00024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX DR WPI; 2001-483447/52.
XX PT Human genome-derived single exon nucleic acid probes useful for analyzing
XX PT gene expression in human foetal liver.
XX PS Claim 27; SEQ ID NO 29989; 639pp + Sequence Listing; English.
XX CC The invention relates to a single exon nucleic acid probe for measuring
XX CC human gene expression in a sample derived from human foetal liver. The
XX CC single exon nucleic acid probes may be used for predicting, measuring and
XX CC displaying gene expression in samples derived from human foetal liver. The
XX CC present sequence is a peptide encoded by a single exon nucleic acid probe
XX CC of the invention. Note: The sequence data for this patent did not form
XX CC part of the printed specification, but was obtained in electronic format
XX CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 560 AA;
Query Match 35.5%; Score 2835; DB 4; Length 560;
Best Local Similarity 99.1%; Pred. No. 4.3e-187;
Matches 559; Conservative 0; Mismatches 1; Indels 4; Gaps 1;
QY 305 TKRKTGSPDRKPSKSKTDNSLSSPLNPKLWCHVHLKKSLSGSPKVKNSKNSKSP 364
Db 1 TKRKTGSPDRKPSKSKTDNSLSSPLNPKLWCHVHLKKSLSGSPKVKNSKNSKSP 60
QY 365 HLEEMKQMSPNKLTNTHFIPKGGPPAKKPGKHSKPLKAKGRSGKILNGQKSTGNSKSP 424
Db 61 HLEEMKQMSPNKLTNTHFIPKGGPPAKKPGKHSKPLKAKGRSGKILNGQKSTGNSKSP 120
QY 425 KGLKATPTKTKMQLTLLDMAKTOKMTRAPNSGCTPTSSKPHKLPALHLIAYKE 484
Db 121 KGLKATPTKTKMQLTLLDMAKTOKMTRAPNSGCTPTSSKPHKLPALHLIAYKE 180
QY 485 NKDRDKRSALSCVISTARLSSSDRLPRLSLVOKRYELLEHKRWASMSBQK 544
Db 181 NKDRDKRSALSCVISTARLSSSDRLPRLSLVOKRYELLEHKRWASMSBQK 240
QY 545 EYLKKEBELKKLKEKAKEREKEMLERLEKQKYEDELTKGNLPAFLVDTPEGLPN 604
Db 241 EYLKKEBELKKLKEKAKEREKEMLERLEKQKYEDELTKGNLPAFLVDTPEGLPN 300
QY 605 TLFGDVAMVFLSCYSGLLLPDAQYPTAVSLMEALSADKGGFLYLNRLVILLOTLQ 664
Db 301 TLFGDVAMVFLSCYSGLLLPDAQYPTAVSLMEALSADKGGFLYLNRLVI----LLQ 356
QY 665 TLLQDEIADYGLGKMLSEIPLTLHVSSELVRLCLRRSDVOESEGSDTDNDKDSAAFE 724
Db 357 TLLQDEIADYGLGKMLSEIPLTLHVSSELVRLCLRRSDVOESEGSDTDNDKDSAAFE 416
QY 725 DNEVDFLEKLETSEFFELTSEEKLIQTALCHRLMTYSVQDHMETRQMSAELWKER 784

Db 417 DNEVDFLEKLETSEFFELTSEEKLIQTALCHRLMTYSVQDHMETRQMSAELWKER 476
QY 785 LAVLKEENDKKAERKQKREMEAKNKENGKGLGKTDKRIIVKFPQVDTEADMLS 844
Db 477 LAVLKEENDKKAERKQKREMEAKNKENGKGLGKTDKRIIVKFPQVDTEADMLS 536
QY 845 AVKSRRLLAIOAKEREIOEREMK 868
Db 537 AVKSRRLLAIOAKEREIOEREMK 560
RESULT 18
AAM30808
ID AAM30808 standard; protein; 560 AA.
XX AC AAM30808;
XX DT 17-OCT-2001 (first entry)
XX DE Peptide #4845 encoded by probe for measuring placental gene expression.
XX KW Probe; microarray; human; placenta; antenatal diagnosis;
XX KW genetic disorder.
XX OS Homo sapiens.
XX PN WO200157272-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US000663.
XX PR 04-FEB-2000; 2000US-0180312P.
XX PR 26-MAY-2000; 2000US-0207456P.
XX PR 30-JUN-2000; 2000US-00608408.
XX PR 03-AUG-2000; 2000US-00632366.
XX PR 21-SEP-2000; 2000US-0234687P.
XX PR 27-SEP-2000; 2000US-0236359P.
XX PR 04-OCT-2000; 2000GB-00024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX DR WPI; 2001-488897/53.
XX PT Human genome-derived single exon nucleic acid probes useful for analyzing
XX PT gene expression in human placenta.
XX PS Claim 27; SEQ ID NO 31077; 654pp; English.
XX CC The present invention relates to single exon nucleic acid probes (SNP:
XX CC see AA13135-AA157546). The present sequence is a peptide encoded by one
XX CC such probe. The probes are useful for producing a microarray for
XX CC predicting, measuring and displaying gene expression in samples derived
XX CC from human placenta. The probes are useful for antenatal diagnosis of
XX CC human genetic disorders
XX SQ Sequence 560 AA;
Query Match 35.5%; Score 2835; DB 4; Length 560;
Best Local Similarity 99.1%; Pred. No. 4.3e-187;
Matches 559; Conservative 0; Mismatches 1; Indels 4; Gaps 1;
QY 305 TKRKTGSPDRKPSKSKTDNSLSSPLNPKLWCHVHLKKSLSGSPKVKNSKNSKSP 364
Db 1 TKRKTGSPDRKPSKSKTDNSLSSPLNPKLWCHVHLKKSLSGSPKVKNSKNSKSP 60
QY 365 HLEEMKQMSPNKLTNTHFIPKGGPPAKKPGKHSKPLKAKGRSGKILNGQKSTGNSKSP 424
Db 61 HLEEMKQMSPNKLTNTHFIPKGGPPAKKPGKHSKPLKAKGRSGKILNGQKSTGNSKSP 120

```
QY 425 KKGLTPKTKMKQMTLLDMAKGTQKMTAPRNSGGTPTSSKPKHKLPPAALHLIAYYKE 484
DB |||||
QY 121 KKGLTPKTKMKQMTLLDMAKGTQKMTAPRNSGGTPTSSKPKHKLPPAALHLIAYYKE 180
DB |||||
QY 485 NKREDKRSALSCVISTARLLSSDDRARLPPELRSLVQKRYELHKKRWASMSSEQRK 544
DB |||||
QY 181 NKREDKRSALSCVISTARLLSSDDRARLPPELRSLVQKRYELHKKRWASMSSEQRK 240
QY 545 EYLKKREELKKLKEKAKERREKEMLERLEKQKRYEDQELTGKLPAPFLVDTPEGLPN 604
DB 241 EYLKKREELKKLKEKAKERREKEMLERLEKQKRYEDQELTGKLPAPFLVDTPEGLPN 300
QY 605 TLFGDVAMVVEFLSCYSGLLLPDAQYPTITAVSLMEALSADKGGFLYLNRLVILLQTLQ 664
DB 301 TLFGDVAMVVEFLSCYSGLLLPDAQYPTITAVSLMEALSADKGGFLYLNRLVILLQTLQ 356
QY 665 TLQDEIAEDYGEIGMKLSEIPLTLHVSSELVRLCLRRSDVOESEGSDTDNKSAAFE 724
DB 357 TLQDEIAEDYGEIGMKLSEIPLTLHVSSELVRLCLRRSDVOESEGSDTDNKSAAFE 416
QY 725 DNEVDFLEKLETSEFFELTSEEKQLIILTALCHRLMTYSVQDHMETRQMSABLWKER 784
DB 417 DNEVDFLEKLETSEFFELTSEEKQLIILTALCHRLMTYSVQDHMETRQMSABLWKER 476
QY 785 LAVLKEENDKKRAEKQKREKEMAKKNGKVENGLGKTDKRIIVKFPQVDTEADMTS 844
DB 477 LAVLKEENDKKRAEKQKREKEMAKKNGKVENGLGKTDKRIIVKFPQVDTEADMTS 536
QY 845 AVKSRRLAIQAKKEREIQEREMK 868
DB 537 AVKSRRLAIQAKKEREIQEREMK 560
RESULT 19
ABB32101
ID ABB32101 standard; peptide; 560 AA.
XX AC ABB32101;
XX DT 01-FEB-2002 (first entry)
XX DE Peptide #4752 encoded by breast cell single exon nucleic acid probe.
XX KW Human; microarray; single exon probe; gene expression; breast; disease;
XX KW Cancer.
XX OS Homo sapiens.
XX PN WO200157271-A2.
XX XX 09-AUG-2001.
XX PD 30-JAN-2001; 2001WO-US0000662.
XX PF 04-FEB-2000; 2000US-0180312P.
XX PR 26-MAY-2000; 2000US-0207456P.
XX PR 30-JUN-2000; 2000US-00608408.
XX PR 03-AUG-2000; 2000US-00632366.
XX PR 21-SEP-2000; 2000US-0234687P.
XX PR 27-SEP-2000; 2000US-0236359P.
XX PR 04-OCT-2000; 2000GB-00024263.
XX XX (MOLE-) MOLECULAR DYNAMICS INC.
XX PA Penn SG, Hanzel DK, Chen W, Rank DR;
XX PI WPI; 2001-496933/54.
XX DR
XX PT New spatially-addressable set of single exon nucleic acid probes, useful
XX PT for measuring gene expression in sample derived from human breast,
XX PT comprises number of single exon nucleic acid probes.
XX XX
XX PS Claim 27; SEQ ID NO 15069; 327pp + Sequence Listing; English.
```

```
XX The invention relates to a spatially-addressable set of single exon
CC nucleic acid probes for measuring gene expression in a sample derived
CC from human breast and BT 474 cells. The method involves contacting the
CC probes with a collection of detectably labelled nucleic acids derived
CC from mRNA of human breast, and then measuring the label bound to each
CC probe of the microarray. The probes are useful for verifying the
CC expression of regions of genomic DNA predicted to encode proteins. They
CC are useful for gene discovery, and for determining predisposition and/or
CC prognosing breast disease. Gene expression analysis is useful for
CC assessing the toxicity of chemical agents on cells. The microarray of
CC this invention presents a far greater diversity of probes for measuring
CC gene expression, with far less bias than expressed sequence tag
CC microarrays. The method is suitable for rapid production of functional
CC information from genomic sequence. The present sequence is a peptide
CC encoded by a single exon nucleic acid probe of the invention. Note: The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 560 AA;
Query Match 35.5%; Score 2835; DB 4; Length 560;
Best Local Similarity 99.1%; Pred. No. 4.3e-187;
Matches 559; Conservative 0; Mismatches 1; Indels 4; Gaps 1;
QY 305 TKRNTGSPDRKPSKSKTDNSSLSSPLNPKLCHVHLKKSLSGSPKVKNSKNSKSP 364
DB 1 TKRNTGSPDRKPSKSKTDNSSLSSPLNPKLCHVHLKKSLSGSPKVKNSKNSKSP 60
QY 365 HLEEMKQMSPNKLTNFIHPKGPAPAKPGKHSKPLKAKGRSGIINGOKSTGNSKP 424
DB 61 HLEEMKQMSPNKLTNFIHPKGPAPAKPGKHSKPLKAKGRSGIINGOKSTGNSKP 120
QY 425 KKGLTPKTKMKQMTLLDMAKGTQKMTAPRNSGGTPTSSKPKHKLPPAALHLIAYYKE 484
DB 121 KKGLTPKTKMKQMTLLDMAKGTQKMTAPRNSGGTPTSSKPKHKLPPAALHLIAYYKE 180
QY 485 NKREDKRSALSCVISTARLLSSDDRARLPPELRSLVQKRYELHKKRWASMSSEQRK 544
DB 181 NKREDKRSALSCVISTARLLSSDDRARLPPELRSLVQKRYELHKKRWASMSSEQRK 240
QY 545 EYLKKREELKKLKEKAKERREKEMLERLEKQKRYEDQELTGKLPAPFLVDTPEGLPN 604
DB 241 EYLKKREELKKLKEKAKERREKEMLERLEKQKRYEDQELTGKLPAPFLVDTPEGLPN 300
QY 605 TLFGDVAMVVEFLSCYSGLLLPDAQYPTITAVSLMEALSADKGGFLYLNRLVILLQTLQ 664
DB 301 TLFGDVAMVVEFLSCYSGLLLPDAQYPTITAVSLMEALSADKGGFLYLNRLVILLQTLQ 356
QY 665 TLQDEIAEDYGEIGMKLSEIPLTLHVSSELVRLCLRRSDVOESEGSDTDNKSAAFE 724
DB 357 TLQDEIAEDYGEIGMKLSEIPLTLHVSSELVRLCLRRSDVOESEGSDTDNKSAAFE 416
QY 725 DNEVDFLEKLETSEFFELTSEEKQLIILTALCHRLMTYSVQDHMETRQMSABLWKER 784
DB 417 DNEVDFLEKLETSEFFELTSEEKQLIILTALCHRLMTYSVQDHMETRQMSABLWKER 476
QY 785 LAVLKEENDKKRAEKQKREKEMAKKNGKVENGLGKTDKRIIVKFPQVDTEADMTS 844
DB 477 LAVLKEENDKKRAEKQKREKEMAKKNGKVENGLGKTDKRIIVKFPQVDTEADMTS 536
QY 845 AVKSRRLAIQAKKEREIQEREMK 868
DB 537 AVKSRRLAIQAKKEREIQEREMK 560
RESULT 20
ABB22640
ID ABB22640 standard; protein; 560 AA.
XX AC ABB22640;
XX XX
```

DT 23-JAN-2002 (first entry)
XX Protein #4639 encoded by probe for measuring heart cell gene expression.
DE Human; gene expression; heart; microarray; vascular system;
KW cardiovascular disease; hypertension; cardiac arrhythmia;
KW congenital heart disease.
XX Homo sapiens.
OS WO200157274-A2.
XX 09-AUG-2001.
XX 30-JAN-2001; 2001WO-US000666.
XX 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
PA Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488999/53.
XX Single exon nucleic acid probes for analyzing gene expression in human
PT hearts.
XX Claim 15; SEQ ID NO 24410; 530pp; English.
XX The present invention relates to single exon nucleic acid probes for
CC measuring human gene expression in a sample derived from human heart (see
CC ABA21535-ABA41305). The present sequence is a protein encoded by one such
CC probe. The probes may be used for predicting, measuring and displaying
CC gene expression in samples derived from the human heart via microarrays.
CC By measuring gene expression, the probes are useful for predicting,
CC diagnosing, grading, staging, monitoring and prognosing diseases of the
CC human heart and vascular system e.g. cardiovascular disease,
CC hypertension, cardiac arrhythmias and congenital heart disease. Note: The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences
XX Sequence 560 AA;
SQ
Query Match 35.5%; Score 2835; DB 4; Length 560;
Best Local Similarity 99.1%; Pred. No. 4.3e-187;
Matches 559; Conservative 0; Mismatches 1; Indels 4; Gaps 1;
QY 305 TKRNTGSPDRKPSKSTDNSSPLNPKLWCHVHLKSLGSPKLVKNSKNSKSP 364
DB 1 TKRNTGSPDRKPSKSTDNSSPLNPKLWCHVHLKSLGSPKLVKNSKNSKSP 60
QY 365 HLEEMKMWSPNKLNTNPHI PKGPPAKPGKHSKPLKAKGRSGILNGKSTGNSKSP 424
DB 61 HLEEMKMWSPNKLNTNPHI PKGPPAKPGKHSKPLKAKGRSGILNGKSTGNSKSP 120
QY 425 KGLKTPTKMOMTLLDMAKTOQWTRAPRNSGGTPTSSKPHLPPAALHLIAYKE 484
DB 121 KGLKTPTKMOMTLLDMAKTOQWTRAPRNSGGTPTSSKPHLPPAALHLIAYKE 180
QY 485 NKREDKRSALSCVISTKARLSSSDRLPPELRSVQKRYELHKKRWASMSSEQRK 544
DB 181 NKREDKRSALSCVISTKARLSSSDRLPPELRSVQKRYELHKKRWASMSSEQRK 240
QY 545 EYLKKREELKKLEKAKERREKEMLEKQKRYEDELTKGNLPAPFLVDTPEGLPN 604
DB 241 EYLKKREELKKLEKAKERREKEMLEKQKRYEDELTKGNLPAPFLVDTPEGLPN 300

QY 605 TLFGDVAMVVEFLSCYSGLLLPDAQYPITAVSLMEALSADKGGFLYLNRLVILLQ 664
DB 301 TLFGDVAMVVEFLSCYSGLLLPDAQYPITAVSLMEALSADKGGFLYLNRLV 356
QY 665 TLLODEIADYGEELGKMLSEIPLTHSVSELVRLCLRRSDVQESGSDTDNKSAAFE 724
DB 357 TLLODEIADYGEELGKMLSEIPLTHSVSELVRLCLRRSDVQESGSDTDNKSAAFE 416
QY 725 DNEVODEFLEKLETSEFFELTSEKLIQILTALCHRLIMTYSVQDHMETROQMSAEIWKER 784
DB 417 DNEVODEFLEKLETSEFFELTSEKLIQILTALCHRLIMTYSVQDHMETROQMSAEIWKER 476
QY 785 LAVLKEENDKKRAEKQKREMEAKNKENGKVENGLGKTDRKKRIKVFEPQVDTAEADMIS 844
DB 477 LAVLKEENDKKRAEKQKREMEAKNKENGKVENGLGKTDRKKRIKVFEPQVDTAEADMIS 536
QY 845 AVKSRLLAIQAQKEREIQEREMK 868
DB 537 AVKSRLLAIQAQKEREIQEREMK 560
RESULT 21
AAM70484
ID AAM70484 standard; protein; 560 AA.
XX AC AAM70484;
XX 06-NOV-2001 (first entry)
XX Human bone marrow expressed probe encoded protein SEQ ID NO: 30790.
DE Human; bone marrow expressed exon; gene expression analysis; probe;
KW microarray; cancer; leukaemia; lymphoma; myeloma.
XX Homo sapiens.
OS WO200157276-A2.
XX 09-AUG-2001.
XX 30-JAN-2001; 2001WO-US000666.
XX 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
PA Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488900/53.
XX Human genome-derived single exon nucleic acid probes useful for analyzing
PT gene expression in human bone marrow.
XX Example 4; SEQ ID NO 30790; 658pp + Sequence Listing; English.
XX The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC bone marrow. They can be used to measure gene expression in bone marrow
CC samples, which may enable the improved diagnosis and treatment of cancers
CC such as lymphoma, leukaemia and myeloma. The present sequence is a
CC protein encoded by one of the probes of the invention
XX Sequence 560 AA;
SQ
Query Match 35.5%; Score 2835; DB 4; Length 560;
Best Local Similarity 99.1%; Pred. No. 4.3e-187;

Matches 559; Conservative 0; Mismatches 1; Indels 4; Gaps 1;	
QY	305 TKRNTGSPDRKPSKSKTDNSSLSPLNPKLWCHVHLKKSLSGSPKVKNSKNSKSP 364
Db	1 TKRNTGSPDRKPSKSKTDNSSLSPLNPKLWCHVHLKKSLSGSPKVKNSKNSKSP 60
QY	365 HLEEMKMSPNKLTNPHIPKGGPPAKPGKHSKPLKAKGRSGILNGKSTGNSKSP 424
Db	61 HLEEMKMSPNKLTNPHIPKGGPPAKPGKHSKPLKAKGRSGILNGKSTGNSKSP 120
QY	425 KKGLTPKTKMKQMTLLDMAKGTQMTAPRNSGGTPTTSKPKHKLPPAALHLIAYYKE 484
Db	121 KKGLTPKTKMKQMTLLDMAKGTQMTAPRNSGGTPTTSKPKHKLPPAALHLIAYYKE 180
QY	485 NKDREDKRSALSCVISTARLLSSSDRALRPEELRSVQKRYEDELTKGKLPAPRLVDTPEGLPN 544
Db	181 NKDREDKRSALSCVISTARLLSSSDRALRPEELRSVQKRYEDELTKGKLPAPRLVDTPEGLPN 300
QY	605 TLFGDVAMVVEFLSCYSGLLLPDAQYPIITAVSLMEALSADKGGFLYLNRLVLIQTLLQ 664
Db	301 TLFGDVAMVVEFLSCYSGLLLPDAQYPIITAVSLMEALSADKGGFLYLNRLVLIQTLLQ 356
QY	665 TLLQDEIAEDYGELGKMLSEIPLTHSVSELVRLCLRRSDVQEESEGSDTDDNKDSAAFE 724
Db	357 TLLQDEIAEDYGELGKMLSEIPLTHSVSELVRLCLRRSDVQEESEGSDTDDNKDSAAFE 416
QY	725 DNEVQDEFLEKLETSEFFELTSEEKQLITLALCHRLMTYSVQDHMETROQMSAELWKER 784
Db	417 DNEVQDEFLEKLETSEFFELTSEEKQLITLALCHRLMTYSVQDHMETROQMSAELWKER 476
QY	785 LAVLKEENDKKRAEKQKKEAKNKENGKVGKGLTKDRKKRIVKFPQVDTEADMTS 844
Db	477 LAVLKEENDKKRAEKQKKEAKNKENGKVGKGLTKDRKKRIVKFPQVDTEADMTS 536
QY	845 AVKSRRLAIQAKKEREIQEREMK 868
Db	537 AVKSRRLAIQAKKEREIQEREMK 560
RESULT 22	
AAM58044	
ID	AAM58044 standard; protein; 560 AA.
XX	AC AAM58044;
XX	AC AAM58044;
XX	05-NOV-2001 (first entry)
XX	Human brain expressed single exon probe encoded protein SEQ ID NO: 30149.
DE	Human; brain expressed exon; gene expression analysis; probe; microarray;
KW	Alzheimer's disease; multiple sclerosis; schizophrenia; epilepsy; cancer.
XX	OS Homo sapiens.
XX	PN WO200157275-A2.
XX	PD 09-AUG-2001.
XX	PF 30-JAN-2001; 2001WO-US0000667.
XX	PR 04-FEB-2000; 2000US-0180312P.
PR	26-MAY-2000; 2000US-0207456P.
PR	30-JUN-2000; 2000US-00608408.
PR	03-AUG-2000; 2000US-00632366.
PR	21-SEP-2000; 2000US-0234687P.
PR	27-SEP-2000; 2000US-0236359P.
PR	04-OCT-2000; 2000GB-00024263.
XX	(MOLE-) MOLECULAR DYNAMICS INC.
PA	
Penn SG, Hanzel DK, Chen W, Rank DR;	
WPI; 2001-483446/52.	
Single exon nucleic acid probes for analyzing gene expression in human brains.	
Example 4; SEQ ID NO 30149; 650pp + Sequence Listing; English.	
The present invention provides a number of single exon nucleic acid probes which are derived from genomic sequences expressed in the human brain. They can be used to measure gene expression in brain cell samples, which may enable the diagnosis and improved treatment of nervous system diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia, epilepsy and cancers. The present sequence is a protein encoded by one of the probes of the invention	
Sequence 560 AA;	
Query Match 35.5%; Score 2835; DB 4; Length 560;	
Best Local Similarity 99.1%; Pred. No. 4.3e-187;	
Matches 559; Conservative 0; Mismatches 1; Indels 4; Gaps 1;	
QY	305 TKRNTGSPDRKPSKSKTDNSSLSPLNPKLWCHVHLKKSLSGSPKVKNSKNSKSP 364
Db	1 TKRNTGSPDRKPSKSKTDNSSLSPLNPKLWCHVHLKKSLSGSPKVKNSKNSKSP 60
QY	365 HLEEMKMSPNKLTNPHIPKGGPPAKPGKHSKPLKAKGRSGILNGKSTGNSKSP 424
Db	61 HLEEMKMSPNKLTNPHIPKGGPPAKPGKHSKPLKAKGRSGILNGKSTGNSKSP 120
QY	425 KKGLTPKTKMKQMTLLDMAKGTQMTAPRNSGGTPTTSKPKHKLPPAALHLIAYYKE 484
Db	121 KKGLTPKTKMKQMTLLDMAKGTQMTAPRNSGGTPTTSKPKHKLPPAALHLIAYYKE 180
QY	485 NKDREDKRSALSCVISTARLLSSSDRALRPEELRSVQKRYEDELTKGKLPAPRLVDTPEGLPN 544
Db	181 NKDREDKRSALSCVISTARLLSSSDRALRPEELRSVQKRYEDELTKGKLPAPRLVDTPEGLPN 240
QY	545 EYLKKREELKKLKEKAKERREKEMLERLEKQRYEDELTKGKLPAPRLVDTPEGLPN 604
Db	241 EYLKKREELKKLKEKAKERREKEMLERLEKQRYEDELTKGKLPAPRLVDTPEGLPN 300
QY	605 TLFGDVAMVVEFLSCYSGLLLPDAQYPIITAVSLMEALSADKGGFLYLNRLVLIQTLLQ 664
Db	301 TLFGDVAMVVEFLSCYSGLLLPDAQYPIITAVSLMEALSADKGGFLYLNRLVLIQTLLQ 356
QY	665 TLLQDEIAEDYGELGKMLSEIPLTHSVSELVRLCLRRSDVQEESEGSDTDDNKDSAAFE 724
Db	357 TLLQDEIAEDYGELGKMLSEIPLTHSVSELVRLCLRRSDVQEESEGSDTDDNKDSAAFE 416
QY	725 DNEVQDEFLEKLETSEFFELTSEEKQLITLALCHRLMTYSVQDHMETROQMSAELWKER 784
Db	417 DNEVQDEFLEKLETSEFFELTSEEKQLITLALCHRLMTYSVQDHMETROQMSAELWKER 476
QY	785 LAVLKEENDKKRAEKQKKEAKNKENGKVGKGLTKDRKKRIVKFPQVDTEADMTS 844
Db	477 LAVLKEENDKKRAEKQKKEAKNKENGKVGKGLTKDRKKRIVKFPQVDTEADMTS 536
QY	845 AVKSRRLAIQAKKEREIQEREMK 868
Db	537 AVKSRRLAIQAKKEREIQEREMK 560
RESULT 23	
AAM05928	
ID	AAM05928 standard; protein; 560 AA.
XX	AC AAM05928;
XX	09-OCT-2001 (first entry)
XX	

DE Peptide #4610 encoded by probe for measuring breast gene expression.
 XX Probe; human; breast disease; breast cancer; development disorder;
 KW inflammatory disease; proliferative breast disease; non-carcinoma tumour.
 XX Homo sapiens.
 XX WO200157270-A2.
 XX 09-AUG-2001.
 XX 29-JAN-2001; 2001WO-US000661.
 XX 04-FEB-2000; 2000US-0180312P.
 XX 26-MAY-2000; 2000US-0207456P.
 XX 30-JUN-2000; 2000US-00608408.
 XX 03-AUG-2000; 2000US-00632366.
 XX 21-SEP-2000; 2000US-0234687P.
 XX 27-SEP-2000; 2000US-0236359P.
 XX 04-OCT-2000; 2000GB-00024263.
 XX (MOLE-) MOLECULAR DYNAMICS INC.
 XX PA Penn SG, Hanzel DK, Chen W, Rank DR;
 XX PI WPI; 2001-476286/51.
 XX DR Novel single exon nucleic acid probe used to measuring gene expression in
 XX a human breast.
 XX PT Claim 27; SEQ ID NO 14668; 322pp; English.
 XX PS The present invention relates to novel single exon nucleic acid probes
 CC (see A100010-AA110067). The present sequence is a peptide encoded by one
 CC such probe. The probes are useful for measuring human gene expression in
 CC a human breast sample, where the probe hybridises at high stringency to a
 CC nucleic acid expressed in the human breast. The probes are useful for
 CC predicting, diagnosing, grading, staging, monitoring and prognosing
 CC diseases of the human breast, particularly those diseases with polygenic
 CC aetiology. The diseases include: breast cancer, disorders of development,
 CC inflammatory diseases of the breast, fibrocystic changes, proliferative
 CC breast disease and non-carcinoma tumours. Note: The sequence data for
 CC this patent did not form part of the printed specification, but was
 CC obtained in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX SQ Sequence 560 AA;

Query Match 35.5%; Score 2835; DB 4; Length 560;
 Best Local Similarity 99.1%; Pred. No. 4.3e-187;
 Matches 559; Conservative 0; Mismatches 1; Indels 4; Gaps 1;

QY 305 TKRKTGSPDRKPSKSKTDNSLSPLNPKLWCHVHLKSLGSLPKVQNSKNSKSP 364
 DB 1 TKRKTGSPDRKPSKSKTDNSLSPLNPKLWCHVHLKSLGSLPKVQNSKNSKSP 60

QY 365 HLEEMKQWSPNKLHTNFHPPKGPAPKPKGKHSKPLKAKGRSKGLNGKSTGNSKSP 424
 DB 61 HLEEMKQWSPNKLHTNFHPPKGPAPKPKGKHSKPLKAKGRSKGLNGKSTGNSKSP 120

QY 425 KKLGLTPKTKMQLLDMAKTQKTRAPRNSGGTPTSSKPKHLPPAALHLIAYYKE 484
 DB 121 KKLGLTPKTKMQLLDMAKTQKTRAPRNSGGTPTSSKPKHLPPAALHLIAYYKE 180

QY 485 NKDREDKRSALSCVTSKTRALSSDRAPLPELRSVLQKRYELLEHKRWASMSSEQRK 544
 DB 181 NKDREDKRSALSCVTSKTRALSSDRAPLPELRSVLQKRYELLEHKRWASMSSEQRK 240

QY 545 EYLKRRRELKKLKEKAKEREKEMLEKQKVEDQELGKNLPAPRLVDTPEGLPN 604
 DB 241 EYLKRRRELKKLKEKAKEREKEMLEKQKVEDQELGKNLPAPRLVDTPEGLPN 300

QY 605 TLFGDVAMVVEFLSCYSGLLLPDAQYPITAVSLMEALSADKGGFLYLNRLVLIQ 664

DB 301 TLFGDVAMVVEFLSCYSGLLLPDAQYPITAVSLMEALSADKGGFLYLNRLVLIQ 356
 QY 665 TLQDEIAEDYGEIGMKLSEIPLTLHSVSELVRLCLRRSDVOESEGSDTDNKSAAPE 724
 DB 357 TLQDEIAEDYGEIGMKLSEIPLTLHSVSELVRLCLRRSDVOESEGSDTDNKSAAPE 416
 QY 725 DNEVQDEFLEKLETSEFFELTSEEKLIQILTALCHRLMTYSVQDHMETRQMSAELWKR 784
 DB 417 DNEVQDEFLEKLETSEFFELTSEEKLIQILTALCHRLMTYSVQDHMETRQMSAELWKR 476
 QY 785 LAVLKEENDKKRAEKOKKEMAKKENGKGVNGKGTDRKKRIVKFEPOVDTEADMS 844
 DB 477 LAVLKEENDKKRAEKOKKEMAKKENGKGVNGKGTDRKKRIVKFEPOVDTEADMS 536
 QY 845 AVKSRLLAIQAKKEREIQEREMK 868
 DB 537 AVKSRLLAIQAKKEREIQEREMK 560

RESULT 24
 ABG40123
 ID ABG40123 standard; peptide; 560 AA.
 XX AC ABG40123;
 XX DT 19-AUG-2002 (first entry)
 XX DE Human peptide encoded by genome-derived single exon probe SEQ ID 29788.
 XX KW Human; single exon probe; asthma; lung cancer; COPD; ILD;
 KW chronic obstructive pulmonary disease; interstitial lung disease;
 KW familial idiopathic pulmonary fibrosis; neurofibromatosis;
 KW tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;
 KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;
 KW pulmonary histiocytosis; lymphangioleiomyomatosis; Kargener syndrome;
 KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
 KW primary ciliary dyskinesia; pulmonary hypertension;
 XX hyaline membrane disease.
 XX OS Homo sapiens.
 XX SS WO200186003-A2.
 XX PN 15-NOV-2001.
 XX PD 30-JAN-2001; 2001WO-US000665.
 XX PF 04-FEB-2000; 2000US-0180312P.
 XX PR 26-MAY-2000; 2000US-0207456P.
 XX PR 30-JUN-2000; 2000US-00608408.
 XX PR 03-AUG-2000; 2000US-00632366.
 XX PR 21-SEP-2000; 2000US-0234687P.
 XX PR 27-SEP-2000; 2000US-0236359P.
 XX PR 04-OCT-2000; 2000GB-00024263.
 XX (MOLE-) MOLECULAR DYNAMICS INC.
 XX PA Penn SG, Hanzel DK, Chen W, Rank DR;
 XX PI WPI; 2002-114183/15.
 XX DR Spatially-addressable set of single exon nucleic acid probes, used to
 XX PT measure gene expression in human lung samples.
 XX PS Claim 27; SEQ ID NO 29788; 634pp; English.
 XX CC The invention relates to a spatially-addressable set of single exon
 CC nucleic acid probes for measuring gene expression in a sample derived
 CC from human lung comprising single exon nucleic acid probes having one of
 CC 12614 nucleic acid sequences mentioned in the specification, or their
 CC complements or the 12387 open reading frames derived from the 12614
 CC probes. Also included are a microarray comprising the novel set of probes

CC ; the novel set of probes which hybridise at high stringency to a nucleic
CC acid expressed in the human lung; measuring gene expression in a sample
CC derived from human lung, comprising (a) contacting the array with a
CC collection of detectably labeled nucleic acids derived from human lung
CC mRNA, and (b) measuring the label detectably bound to each probe of the
CC array; identifying exons in a eukaryotic genome, comprising (a)
CC algorithmically predicting at least one exon from genomic sequences of
CC the eukaryote; and (b) detecting specific hybridisation of detectably
CC labeled nucleic acids from eukaryotic lung mRNA, to a single exon probe,
CC having a fragment identical to the predicted exon, the probe is included
CC in the above mentioned microarray; assigning exons to a single gene,
CC comprising (a) identifying exons from genomic sequence by the method
CC above and (b) measuring the expression of each of the exons in several
CC tissues and/or cell types using hybridisation to a single exon
CC microarrays having a probe with the exon, where a common pattern of
CC expression of the exons in the tissues and/or cell types indicates that
CC the exons should be assigned to a single gene; a peptide comprising one
CC of 12011 sequences, mentioned in the specification, or encoded by the
CC probes/open reading frames (ORF). The probes are used for gene expression
CC analysis, and for identifying exons in a gene, particularly using human
CC lung derived mRNA and for the study of lung diseases such as asthma, lung
CC cancer, chronic obstructive pulmonary disease (COPD), interstitial lung
CC disease (ILD), familial idiopathic pulmonary fibrosis, neurofibromatosis,
CC tuberous sclerosis, Gaucher's disease, Niemann-Pick disease, Hermansky-
CC Pudlak syndrome, sarcoidosis, pulmonary haemosiderosis, pulmonary
CC histiocytosis, lymphangioleiomyomatosis, pulmonary alveolar proteinosis,
CC Karagener syndrome, fibrocystic pulmonary dysplasia, primary ciliary
CC dyskinesia, pulmonary hypertension and hyaline membrane disease. The
CC present sequence is a peptide/protein encoded by a single exon probe of
CC the invention. Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic format
CC directly from WIFO at ftp.wipo.int/pub/published_pct_sequences

XX Sequence 560 AA;

Qy	305	TKRKNVTSPPDRKPSKSKSTNDSLSPLNPKLWCHVHLKSLGSPKVKNSKNSKSP	364
Db	1	TKRKNVTSPPDRKPSKSKSTNDSLSPLNPKLWCHVHLKSLGSPKVKNSKNSKSP	60
Qy	365	HLEEMKMKNSPKLHTNPHIPKGGPPAKPGKHSDPKLKAKGRSGILNGKSTGNSKP	424
Db	61	HLEEMKMKNSPKLHTNPHIPKGGPPAKPGKHSDPKLKAKGRSGILNGKSTGNSKP	120
Qy	425	KKGLTPTKTKMKTLLDMAGTKQWTRAPNSGCTPTSSKPKHLPALHLIAYYKE	484
Db	121	KKGLTPTKTKMKTLLDMAGTKQWTRAPNSGCTPTSSKPKHLPALHLIAYYKE	180
Qy	485	NKREDKKSALSCVTSKTRALLSSSDRALPELRSVLQKRYELHKKRWASMEBQRK	544
Db	181	NKREDKKSALSCVTSKTRALLSSSDRALPELRSVLQKRYELHKKRWASMEBQRK	240
Qy	545	EYLKKKREELKKLEKAKEREKEMLERLEKQKRYEDELTKGNLPAFLVDTPEGLPN	604
Db	241	EYLKKKREELKKLEKAKEREKEMLERLEKQKRYEDELTKGNLPAFLVDTPEGLPN	300
Qy	605	TLFGDVMVVEFLSCYSGLLLPDAQYPTAVSLMEALSGADGGFLYLNRLVILLQTLIQ	664
Db	301	TLFGDVMVVEFLSCYSGLLLPDAQYPTAVSLMEALSGADGGFLYLNRLVILLQTLIQ	356
Qy	665	TLLQDEIAEDYGEKMKLSEIPLTLHVSSELVRLCLRRSDVQEESESDTDNDKDSAAFE	724
Db	357	TLLQDEIAEDYGEKMKLSEIPLTLHVSSELVRLCLRRSDVQEESESDTDNDKDSAAFE	416
Qy	725	DNEVDEFEKLETSSEFFELTSEEKLOILTALCHRLIMTYSVODHMETRQMSAELWKER	784
Db	417	DNEVDEFEKLETSSEFFELTSEEKLOILTALCHRLIMTYSVODHMETRQMSAELWKER	476
Qy	785	LAVLKEENDKKRAEKQKREKEMAKNKENGKVENGLGKTRDKRKRIIVKFFPQVDTEADMIS	844

Db	477	LAVLKEENDKKRAEKQKREKEMAKNKENGKVENGLGKTRDKRKRIIVKFFPQVDTEADMIS	536
Qy	845	AVKSRRLLAIOAKKEREIQEREMK	868
Db	537	AVKSRRLLAIOAKKEREIQEREMK	560
RESULT 25			
AA000760	AA000760 standard; protein; 513 AA.		
XX	AC	AA000760;	
XX	DT	01-OCT-2001 (first entry)	
XX	DE	Human bone marrow protein, SEQ ID NO: 123.	
XX	KW	Human; bone marrow; antiinflammatory; cytostatic; neuroprotective;	
XX	KW	antiviral; antibacterial; antifungal; anti-HIV; haemostatic;	
XX	KW	immunosuppressive; gene therapy; cytokine cell proliferation;	
XX	KW	cell differentiation modulator; immune disorder; infection; cancer;	
XX	KW	human immunodeficiency virus; HIV; autoimmune disorder; haemophilia.	
XX	OS	Homo sapiens.	
XX	PN	WO200153453-A2.	
XX	PD	26-JUL-2001.	
XX	PF	23-DEC-2000; 2000WO-US034960.	
XX	PR	23-DEC-1999; 98US-00471275.	
XX	PR	21-JAN-2000; 2000US-00488725.	
XX	PR	25-APR-2000; 2000US-00523117.	
XX	PR	20-JUN-2000; 2000US-00598042.	
XX	PR	19-JUL-2000; 2000US-00620312.	
XX	PR	03-AUG-2000; 2000US-00653450.	
XX	PR	14-SEP-2000; 2000US-00662191.	
XX	PR	19-OCT-2000; 2000US-00693036.	
XX	PR	30-NOV-2000; 2000US-0250583P.	
XX	PA	(HYSE-) HYSEQ INC.	
XX	PI	Ford JE, Boyle BJ, Tang YT, Liu C, Auandi V, Chen R, Ma Y;	
XX	PI	Ren F, Wang J, Werhman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao Qa;	
XX	PI	Zhou P, Drmanac RT;	
XX	WPI	WPI; 2001-488707/53.	
XX	N-PSDB	N-PSDB; AAH89879.	
XX	PT	Novel bone-marrow-expressed polynucleotides and polypeptides, useful for treating e.g. cancer and immune deficiency disorders.	
XX	PS	Claim 10; Page 292-293; 648pp; English.	
XX	CC	The present sequence is one of 251 novel human polypeptides encoded by a bone marrow-expressed polynucleotide. The polynucleotide and the immune polypeptide encoded by it are useful in the treatment of various immune deficiencies and disorders. The deficiencies and disorders may be genetic, may be caused by a viral (e.g. HIV), bacterial or fungal infection, or may result from an autoimmune disorder, a coagulation disorder (e.g. haemophilia), inhibition of tumour cell proliferation, suppression of an inflammatory response or treatment of a nervous system disorder such as Alzheimer's disease. Detection of the presence or increased expression of the polynucleotide or the protein it encodes is useful for the diagnosis and/or prognosis of one or more types of cancer. The polynucleotide and polypeptide can be used as nutritional sources or supplements and in the screening of chemical compounds as potential drugs	
XX	SQ	Sequence 513 AA;	
Query Match 28.9%; Score 2305.5; DB 4; Length 513;			
Best Local Similarity 63.5%; Pred. No. 1.6e-150;			

Matches 477; Conservative 5; Mismatches 8; Indels 261; Gaps 7;	
QY 762	MTYSVDHMETRQMSAELWKERLAVLKEENDKKRAEKQKRKEMEAKNKENGKVENGLGK 821
Db 1	MTYSVDHMETRQMSAELWKDLRLVLKEENDKKRAEKQKRKEMEAKNKENGKVENGLGK 60
QY 822	TDRKKIIVKPEPOVDTEAEDMISAVKSRRLLAIOAKKERHIOEREMKVKLEROAEEERIR 891
Db 61	TDRKKIIVKPEPOVDTEAEDMISAVKSRRLLAIOAKKERHIOEREMKVKLEROAEEERIR 107
QY 882	HKHAAAEKAFQEGIAKAKLWMRRTPICTDNRHNRWYLFSDVPLFIEKGWVHDSIDYRF 941
Db 108	-----ELSLGF----- 113
QY 942	NHCKDHTVSGDEDYCPRSKANLGNKSNMTOHGTATEVAVETTTPKQGNLWFLCDSQ 1001
Db 114	-----CPYR-----FLCDSQ 123
QY 1002	KELDELLNCLHPQIGRESQIKERLEKRYQDIHSHILARKPNLGLKSCDGNQELLNFLRS 1061
Db 124	KELDELLNCLHPQIGRESQIKERLEKRYQDIHSHILARKPNLGLKSCDGNQELLNFLRS 183
QY 1062	DLIEVATRLQKGLGVVEETSEPEARVISLEKLKDFGECVIALQASVIKKFLOGFNAPKQ 1121
Db 184	DLIEVATRLQKGLGVVEETSEPEAR----- 209
QY 1122	KRRKLOSEDSAKTEEVDEEKKMVEEAKVASALEKWKTAIREAQTFSRMHVLLGMLDACIK 1181
Db 210	-----VASALEKWKTAIREAQTFSRMHVLLGMLDACIK 242
QY 1182	WMSAENARCKVCPKKGDDKILDCENKAFHLFCLRPALYEVPPGEWOCFACOPATAR 1241
Db 243	WMSAENARCKVCRKKGDDKILDCENKAFHLFCLRPALYEVDP----- 288
QY 1242	RNSGRNYYTESASEDESDSEDEEEEEEEDYEVAGLRLRPKRTIRGKHSVIP 1301
Db 289	-----VPRKRTIRGKHSVIP 304
QY 1302	AARSGRRPGKPHSTRSQKAPPVDDAEVDELVLQTKSSRRSQSLELQKCEILLHKIVK 1361
Db 305	AARSGRRPGKPHSTRSQKAPPVDDAEVDELVLQTKSSRRSQSLELQKCEILLHKIVK 364
QY 1362	YRFSWPF-----RBPVTRDEADYDVIITHPMDFTQVQNKCS 1398
Db 365	YRFSWPFRTCLSGRTAVKAVQILHLVLLHREPVTRDEADYDVIITHPMDFTQVQNKCS 424
QY 1399	CGSYRSVQBFLLDMKQVFTNAEYVNCRGSHVLSVMYKTEQCLVLLHKLPGHPYVRRK 1458
Db 425	CGSYRSVQBFLLDMKQVFTNAEYVNCRGSHVLSVMYKTEQCLVALLHKLPGHPYVRRK 484
QY 1459	KKFPDLAEDGDSPEAVGQSRDEDRRSRE 1489
Db 485	KKFPDLAEDGDSPEAVGQSR--GRRQKK 513

RESULT 26
ABR41346
ID ABR41346 standard; protein; 209 AA.
XX ABR41346;
AC ABR41346;
XX
XX
DT 02-JUN-2003 (first entry)
XX Human DITHP transcription factor.
DE
XX Human; dithp; diagnostic and therapeutic polynucleotide; diagnosis;
KW cancer; cell proliferative disorder; autoimmune disorder;
KW inflammatory disorder; infection; hormonal disorder; metabolic disorder;
KW neurological disorder; gastrointestinal disorder; transport disorder;
KW connective tissue disorder; drug screening; proteome analysis;
KW gene therapy; antisense therapy; genotyping; transgenic animal; knock in;
KW disease model; toxicological testing; transcript imaging;
KW transcription factor.

XX OS Homo sapiens.	
XX PN WO200297031-A2.	
XX PD 05-DEC-2002.	
XX PF 27-MAR-2002; 2002WO-US010056.	
XX PR 28-MAR-2001; 2001US-0279619P.	
PR 29-MAR-2001; 2001US-0280067P.	
PR 29-MAR-2001; 2001US-0280068P.	
PR 16-MAY-2001; 2001US-0291280P.	
PR 17-MAY-2001; 2001US-0291829P.	
PR 17-MAY-2001; 2001US-0291849P.	
PR 19-JUN-2001; 2001US-0299428P.	
PR 20-JUN-2001; 2001US-0299776P.	
PR 20-JUN-2001; 2001US-0300001P.	
XX (INCY-) INCYTE GENOMICS INC.	
XX Daffo A. Jones AL, Tran AB, Dahl CR, Gietzen D, Chinn J;	
PI Dufour GE, Hillman JL, Yu JY, Tuason O, Yap PE, Amshay SR;	
PI Daughtery SC, Dam TC, Liu TF, Nguyen DA, Kleefeld Y, Gerstin EH;	
PI Peralta CH, David MH, Lewis SA, Chen AJ, Panzer SR, Harris B;	
PI Flores V, Marwaha R, Lo A, Lan RY, Urashka ME;	
XX WPI: 2003-129518/12.	
DR N-PSDB; ACC46286.	
XX Novel human diagnostic and therapeutic polypeptide useful for identifying	
PT test compound which specifically binds to a polypeptide encoded by human	
PT diagnostic and therapeutic polynucleotide, and to induce antibodies.	
XX Claim 27; SEQ ID NO 881; 591pp; English.	
PS The invention relates to novel human diagnostic and therapeutic	
XX polynucleotides designated dithp (ACC46080-ACC46749) and to their encoded	
CC proteins (DITHP; ABR41136-ABR41812). The invention also relates to	
CC polynucleotide sequences at least 90% identical to the dithp cDNA	
CC sequences of the invention; recombinant vectors, host cells and	
CC transgenic organisms comprising a dithp nucleic acid sequence; the	
CC recombinant production of DITHP proteins; antibodies specific for DITHP	
CC proteins; microarrays comprising dithp nucleic acid sequences; methods of	
CC detecting dithp nucleotide and protein sequences; methods of screening	
CC for compounds which specifically bind a DITHP protein; and methods of	
CC assessing the toxicity of test compounds using a dithp hybridisation	
CC probe. Dithp nucleic acid sequences and DITHP proteins may be used in the	
CC diagnosis of a wide variety of conditions including cancer and other cell	
CC proliferative disorders; autoimmune or inflammatory disorders; bacterial,	
CC viral, fungal or parasitic infections; hormonal disorders; metabolic	
CC disorders; neurological disorders; gastrointestinal disorders; transport	
CC disorders; and connective tissue disorders. They may also be used to	
CC screen for modulators of protein activity or gene expression. DITHP	
CC proteins can additionally be used in analysis of the proteome of a tissue	
CC or cell type and to induce antibodies. The dithp nucleic acids are	
CC additionally useful in somatic or germline gene therapy of the disorders	
CC mentioned above, as a source of antisense sequences, as a source of	
CC probes and primers, in genotyping and identification of individuals, in	
CC the generation of transgenic animal models of human disease or knock in	
CC humanised animals, in toxicological testing, and in transcript imaging.	
CC The present sequence represents a DITHP protein which has transcription	
CC factor activity. Note: The sequence data for this patent did not form	
CC part of the printed specification, but was obtained in electronic format	
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences	
XX Sequence 209 AA;	
SQ Query Match 13.5%; Score 1078; DB 6; Length 209;	
Best Local Similarity 96.7%; Pred. No. 3.5e-66;	
Matches 204; Conservative 2; Mismatches 3; Indels 2; Gaps 1;	

QY 1279 YEVALRLRPRTIRGKHSVIPPAARSGRRPGKPHSTRSQKAPPVDDAEVDELVLQ 1338

Db 606 VLEALVGNDSG--PLCELLFFELTAIFAEEVEEVAKEQLDADTKGCSLAKSLDLD 663
QY 689 LHSVSELVRLCURRSDVOEESGSDT-----DNKSAAPEDNEVQDEFLEKLETSEF 741
Db 664 SCTLSEILRLHLASGADVTSANAKRYQKRGGFATDDACMELRLSNPSLWKLSTSV 723
QY 742 FELTSEELQILTALCHRLTMYVSQD-----HMETQOQMSA 778
Db 724 YDLTFGEKKILHALCGKLLTLVSRDFIEDYDILROAKQBFRELKAEQHRKEBEAAA 783
QY 779 ELWKERLAVLKEENDKRAEKAKEMBAKN-----KENGKVENGLGKTDRKKRIV 829
Db 784 RIRKKEEKLKQEQKMEKQEKLEKDEQRNSTADISIGEEREDFTSISKOTEQK-- 841
QY 830 KPEQVDTAEADMISAVKSRRLATQAK-----KEREIQE-----REMKVKLRQA-E 876
Db 842 ELDQMFTEDEDDPGSHKGR-----RGKRGQNGFKFETRQEQINCVCVTRRELLTADBEALK 897
QY 877 ERIRKHKAAAKAQAQEGIAKAKLVMRRTPIGTDNRHNYMLFSDEVPLFIEK---GMV 933
Db 898 QEHQKKELEK-KTSAIACTNIF---PLGRDRMYRYWIF-PSIFGLFIEDYSGLT 951
QY 934 HDSI-----DYRFNHCKDHTVSGDEDYCPRSKKN-LCKKASMMTQGTATEVAVETTT 987
Db 952 EMLLPSPSSFNQVQSDPQVS-----TKTGEPLMSESTSNIDQG-PRDHSVQLPK 1002
QY 988 PKQGNLWFLCDSQKELDELNCLHPQIGRESQKRL--EK-----RYQDIHSH 1037
Db 1003 PVHKENRWCYSSCEQLDQLEALNSGRHRESALKETLLQEKSRICQAQLARFSE--EKPH 1060
QY 1038 LARKNGLGK-----SCDGNQ-----ELNFLASDLIEVATRLQKGLGVVET- 1081
Db 1061 FSDKPPQPSKPTYSGRSSNAYDPSQMAEQLELRDLRDLIEDRIYQGLGAIKVTYD 1120
QY 1082 -----SEFEA---RVISLEKL-----KDFGECVIALQASVKKFLQGFMAPKQKRR 1124
Db 1121 RHIMSALESRYELLSEENKENGIIKTWNEDVEWEIDEQTKVIVK--DRLLGKITETP 1178
QY 1125 KLOSDSATEVDE-----EKQWVEBAKVAS-----ALEKWTAI 1160
Db 1179 STVSTNASTPQSVSSVWVHYLAMALFQIEQIERRELKAPLADSDSGRSYKTVLDRWBSL 1238
QY 1161 REAQTFRMHVLLGMLDACIKWMAENARCKVCKGEDDKLILCDCKNAFHLFCLRP 1220
Db 1239 LSSASLSQVFLHSLTDRSVIWSKILNARCKICRKGDAENWVLCDCGDRGHHTYCVRP 1298
QY 1221 ALYEVDPGEWQCPACQATARNRSGRNYTESASEDGEDDESDEEE-----EEEEEEEE 1276
Db 1299 KLKTVEGDWFCPCRPQRCRLSFRQRPSPLESDVEDSDSGDEDDVDGDEEGQSEE 1358
QY 1277 EDEYV-----AGLRRLRPRTI-----RGKHSVIPPAARSGRRRPGKPHSTR 1318
Db 1359 EYEVQEDDSDQEEVSLPKRGQPQVRLPVKTRGKLSSSPSRSGQQQEGRYPSRQ 1418
QY 1319 SQPK-----APPVD-----DAEVEL----- 1334
Db 1419 STPKTTVSSKTRSLRKINSAPPTTKSLRIASRSTRSHGFLQADVFVLLSPRRKRG 1478
QY 1335 -----VLPQKSRRSQSL----- 1347
Db 1479 KASANTPNSPNFNFRVIATKSEQSRSVNIASKLSLOESKRCRCKRQSPSPSPVT 1538
QY 1348 -----ELQKCEILHKIVKRYFSMPFPVTRDAEDYDVITHMPDQTVQ 1394
Db 1539 LGRSSGRGGVHELSAFQELVVELVRHDDSWFLKLVSKIQVPDYDIKKPIALNIIR 1598
QY 1395 NKCSGYSRVSQEFITDMQVFTNAEYVNCRS 1427
Db 1599 EKVNKEYKLASEFIDDLIELMFSNCFEYNPRT 1631

ADP12578
ID ADP12578 standard; protein; 1674 AA.
XX
AC ADP12578;
XX
DT 12-AUG-2004 (first entry)
XX
DE Protein encoded by mRNA of the invention #188.
XX
KW transplant rejection; immune system; rheumatoid arthritis; lupus;
inflammatory bowel disease; multiple sclerosis; HIV; AIDS.
XX
OS Homo sapiens.
XX
PN WO2004042346-A2.
XX
PD 21-MAY-2004.
XX
PF 24-APR-2003; 2003WO-US012946.
XX
PR 24-APR-2002; 2002US-00131831.
PR 20-DEC-2002; 2002US-00325899.
XX
PA (EXPR-) EXPRESSION DIAGNOSTICS INC.
XX
PI Wohlgemuth J, Fry K, Woodward R, Ly N, Prentice J, Morris M;
PI Rosenberg S;
XX
WI 2004-400724/37.
XX
PT Diagnosing or monitoring transplant rejection, e.g. heart, kidney, liver,
pancreas, pancreatic islet, lung, bone marrow or stem cell transplant
rejection, in an individual, comprises detecting the expression level of
the genes.
XX
PS Claim 65; SEQ ID NO 2587; 1762pp; English.
XX
CC The present invention relates to diagnosing or monitoring transplant
rejection, e.g. cardiac or kidney transplant rejection, in an individual
comprises detecting the expression level of one or more genes. The
methods, system and kits are useful in diagnosing or monitoring
transplant rejection, e.g. heart, kidney, liver, pancreas, pancreatic
islet, lung, bone marrow or stem cell transplant rejection,
xenotransplant rejection or mechanical organ replacement rejection, in an
individual. The method is also useful in assessing the immune status of
an individual. The methods are also useful in diagnosing and monitoring
diseases that involve the immune system, e.g. rheumatoid arthritis,
lupus, inflammatory bowel diseases, multiple sclerosis, HIV/AIDS or
viral, bacterial or fungal infection. The present sequence represents a
protein that is encoded by the mRNA of the invention.
XX
SQ Sequence 1674 AA;
Query Match 11.7%; Score 930.5; DB 8; Length 1674;
Best Local Similarity 23.2%; Pred. No. 1.1e-54;
Matches 397; Conservative 263; Mismatches 532; Indels 521; Gaps 67;
QY 3 PLLGRKPF-----PLVNPLPGSEPPFTTPTQOAFRTREEYEARLERYERWTCKSTGSS 58
Db 152 PLLHRKPFVRQKPPADLRPDEVFY-CKVTNEIFRHYDDFFERTILCSLVMSCAVTRP 210
QY 59 QLTKEAMEEVEQVAELLKESFPAYEKLVLVEMVH-----HNTASLEKLVDTAWLEIMTKY 114
Db 211 GLTYQEALESKKARQNL-QSFP---EPLIIPVLYLTSLTRSRLUHEICDDIFAVKORY 266
QY 115 AVGECDEFEVGEKMLKVKIKVHIHPLEKVDDEATEKSGDGDSPSSDKSNSQIAQDHQ 174
Db 267 FVEETVEVIRNNGARLQCTILEVLP-----PS-----HQNGFANGHV 303
QY 175 KK---ETVVKDEGRRESINDRARRSPKLTSLKKGKRWAPPKFLPHKVDV----KLQ 227
Db 304 NSVDGTTIISDSDSETQS-----CSFQNGKKKDAIDPLL-FKYKVQPTKKEL 351

QY 59 QLTKEAMEERQEVALLKEBPAYEYKLVLEWVH-----HNTASLEKLVDTAWLEIMTKY 114
DB 61 GLTYQEALESEKARQNL-QSGF---EPLIIPVLYLTSLTHRSRLHEICDDIFAFVXKDRY 116
QY 115 AVGECDEPVGKEMKLVKVIKIHLEKVEDEATEKSGDAGCDSPSSDKENSSQIAQDHQ 174
DB 117 FVEETVEVRNNGARLQCRILEVLP-----PS-----HQGFANGHV 153
QY 175 KK---ETVVVEDEGRRESINDRARRSPKLPSTLSKKGERKWAAPPKFLPHKYDV-----KLQ 227
DB 154 NSVDGETIISDSDDSETQS-----CSFQNGKKDAIDPLL-FYKVKVQPTKKEL 201
QY 228 NEDKIIISNPADSLRTERPPNKEIVRYFIRHNLRACTGENAPVWVEDELVKKSLPS- 286
DB 202 HESAI---VKATQISRRKHLFSROKLKFLKQHC-----BPQDGVIK---IKASSUSTY 249
QY 287 -----KFSDFLLDPKYMTLPSTPKRNTGSPDRKPSKK---SKTDNSLSPLNKLWC 338
DB 250 KIAEQDFSYFFPDPTTIFSPANRRG-----RPPKRIHSQEDN----- 290
QY 339 HVHLKKSISGSPKLVKNSKNKSPBEHLEEMMMKMSPNKLJHTNFHIPPKGPPAKPKGKHS 398
DB 291 -VANKQTLA-----SYRSKATKER----- 308
QY 399 DKPLKAKGRSGILNGQSTGNSKSPKGLKTPKTKMKQMTLLDMAKGTQKMTAPRNSG 458
DB 309 DKLLK-----QSEMSLAPEKAKLKR----- 329
QY 459 GTPRTSSKPHKLPPAALHLIAYKENKDREDKRSALSCVISTARTLLSSEDRARLPEEL 518
DB 330 -----EKADALEAKKEKEDKEKR-----BEL 352
QY 519 RSLVQKRYELLEKKRWASMSQEKYLLKKREBLKKLKEKAKERKEKEMLERLEKOK 578
DB 353 KKIVEEE-----RLKKKEBERLKVEREKEREKLRB-EKRKYVEYLKQWSKPR 399
QY 579 RYEDDELGT-KULPAFLRVDPTEGIPNTLFGDVAMVVEFLSCYSGLLLPDAQYP---ITAV 635
DB 400 ---EDNECDLDELPEPTPVKT--RLPPEIFGDALMVLEFLNAFGLFDLQDEFPDGVTL 455
QY 636 SLMEAL-SADKGGFLYLNRLVILLQTLLOLDE----- 670
DB 456 VLEALVGNDSG--PLCELLFFFTAIPTAQIAEEEEEVEAKEQLTDATDKLTEALDEDA 513
QY 671 -----IADYGEL--GMKLSIPLTHSVSELVRLCLRRSDVQSESGSDTD- 715
DB 514 DPTKSALSAVASLAAAWPOLHQGCSLSLDLDSCTLSLRLHLILASGADVTSANAKRY 573
QY 716 -----DNKDSAAFEDNEVQDEFLKLETSSEFELTSEKLOILTALCHRLIMTYSVOD- 768
DB 574 QKRGGFATDDACMELRLSNPSLVKLGJSTSVYDITPGEKMLHALCGKLLTVLSTRDF 633
QY 769 -----HMETRQMSAELMWERLAVLKEENDKKRAEKQKREKEME 806
DB 634 IEDYVDILLQAQOEFRELKAEQHRKEBEAAAIRKKEELKQEQKQKQKELKEDB 693
QY 807 AKNGKGVENGKLTDRKRIKVPQVDTEA-----EDMI-----SAVKSRLALAI 854
DB 694 ---QRNSTADISIGEEREDFTSIESK-DTEQKELDQDMVTDEDDPGSHKRGRRGKR 749
QY 855 Q-AKVERIOEREMKVLE--RQAEERIRKHAARAEKAFQEGIAKALVWRRTIGCTR 911
DB 750 QNGFKEFTROEQINCVTREPLTADDEBALQEHQREKELEKIQSAIACTNIPFLGRDR 809
QY 912 NHRNRYLTSDEVPGLPIEK---GWHDISI-----DYRFNHCKDHTVSGDEDCPSRKA 963
DB 810 MYRRWIP-PSIPGLFIEDYSGLTEDMLPRPSSFQNNVQSDPQVS-----TKTG 860
QY 964 N-LGKNASNTQGTATEVAVETTTPKQGNLWFLCDSQKELDELLNCLHPQIGRESQK 1022
DB 861 EPLMSESTSNIDQG-PRDHSVQLPKPVHKNPNCWCFSSCEQLDQLTALNSRGHRESALK 919

QY 1023 ERL--EK-----RYODIIHSTHARKPNLGLK-----SCDGNQ-----ELLN 1057
DB 920 ETLQEKSRICAQLARSP--EKHFSDKPODPKPYTSRGRSSNAYDPSQWCAEKQLEL 977
QY 1058 FLRSLIEAVTRLOKGLGYVEET-----SEFPA---RVISLEKL-----KQFGE 1099
DB 978 RLROFLLDIEDRIYQGTGAIKVTDRHWRSALESGRYELLSEENKENGIIKTVNEEDVEE 1037
QY 1100 CVIALQASVIKFLQGFMAPKQKRLQSEDSAKTEEVE-----EKKM 1143
DB 1038 MEIDEQTKVIK--DRLLGIKTETPTSTVSTNASTPQSVSSVVHYLAMALFOIEQIERRF 1095
QY 1144 VEEAKVAS-----ALEKWKTAIRBAQTSRSMHVLGMLDACIKWDMSAENARCKVCP 1195
DB 1096 LKAPLDASDSGRSVKTVLDRWRESLSASLSQVFLHLSTLDRSVINWSKSLNARCKICR 1155
QY 1196 KKGEDDKLILCDECNKAFHLCLRPALYEPDGEWQCPACQAPATARRNSRGRNYYTESAS 1255
DB 1156 KKGDAENMWLDCDGRGHHTYCVRPKLTVPEDGWFCECPKQKRSRLSSRQRPSPLESD 1215
QY 1256 EDSDEDSDEEE-----EEEEEEEDYEVA-----GLRLRPKTIIRGKHSVIPPAARSGR 1307
DB 1216 EDVEDSGEGDEVDGDEEGQSEEEYEVEQDEDDGRL-EVKT-RGKLSSSFSGRGOQQ 1273
QY 1308 RPKKKPHSTRSOPK-----APVD-----DAEV 1331
DB 1274 EPGRYPSRQOSTPTKTVSSKTGRSLRKINSAPPTETYSKLRIASRSTPHSHGLOADVTV 1333
QY 1332 DEL-----VLQTKRSTRRQSL----- 1347
DB 1334 ELLSPRRKRRGRKSANNTPENSPPNPRTVATKSSQSRSVNTASKLSLQESSESKRCR 1393
QY 1348 -----ELQCEILHKIVKRYFSWPFPREPVTDEAEDYVDV 1383
DB 1394 KRQSPSPVTLGRSSRQGGVHLSAFEQLVVELVRHDDSWPFLKLVSQIQVPDYDI 1453
QY 1384 ITHMDFTQVONKCSGYSRSVQBFPLTDMKOVFTNAEVYNCRG 1427
DB 1454 IKKPIALNIIREKVNKCEYKLASEFIDIELMFNSCFEYNPNT 1497
RESULT 31
AAU16181
ID AAU16181 standard; protein; 141 AA.
XX
AC AAU16181;
DT
XX
XX 07-NOV-2001 (first entry)
XX Human novel secreted protein, Seq ID 1134.
DE
XX
XX Human; immunosuppressive; antiarthritic; antirheumatic; cytostatic;
KW cardiant; vasotropic; cerebroprotective; nootropic; neuroprotective;
KW antibacterial; virucide; fungicide; ophthalmological; vulnery;
KW secreted protein; rheumatoid arthritis; hyperproliferative disorder;
KW cerebrovascular disorder; cardiac arrest; cerebrovascular disorder;
KW cerebral ischaemia; angiogenesis; nervous system disorder;
KW Alzheimer's disease; infection; ocular disorder; corneal infection;
KW wound healing; epithelial cell proliferation; skin ageing; food additive;
KW preservative; antiproliferative.
XX
OS Homo sapiens.
XX
XX WO200155322-A2.
XX
XX
XX 02-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US001341.
XX
XX 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.

PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0234999P.
PR 27-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0248474P.
PR 08-NOV-2000; 2000US-0248475P.
PR 08-NOV-2000; 2000US-0248476P.
PR 08-NOV-2000; 2000US-0248477P.
PR 08-NOV-2000; 2000US-0248478P.
PR 08-NOV-2000; 2000US-0248523P.
PR 08-NOV-2000; 2000US-0248524P.
PR 08-NOV-2000; 2000US-0248525P.
PR 08-NOV-2000; 2000US-0248526P.
PR 08-NOV-2000; 2000US-0248527P.
PR 08-NOV-2000; 2000US-0248528P.
PR 08-NOV-2000; 2000US-0248532P.
PR 08-NOV-2000; 2000US-0248609P.
PR 08-NOV-2000; 2000US-0248610P.
PR 08-NOV-2000; 2000US-0248611P.
PR 08-NOV-2000; 2000US-0248613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249246P.
PR 17-NOV-2000; 2000US-0249255P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 11-DEC-2000; 2000US-0251990P.
PR 05-JAN-2001; 2001US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-488783/53.
DR N-PSDB; AAS26168.
XX
PT New nucleic acid molecules encoding 461 human secreted proteins for
PT diagnosing, preventing, treating or ameliorating medical conditions and
PT used as food additives or preservatives.
XX
PS Claim 11; SEQ ID NO 1134; 980pp; English.
XX

CC The invention relates to isolated nucleic acid molecules and their
CC encoded secreted proteins. The nucleic acids and proteins are used to
CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,
CC rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used
CC in diagnosing a pathological condition or susceptibility to a
CC pathological condition. Antibodies to the proteins can also be used in
CC alleviating symptoms associated with the disorders and in diagnostic
CC immunoassays e.g. radioimmunoassays or enzyme linked immunosorbant assays
CC (ELISA). Disorders which are diagnosed or treated include autoimmune
CC diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g.
CC neoplasms e.g. of the breast or liver, cardiovascular disorders e.g. cardiac
CC arrest, cerebrovascular disorders e.g. cerebral ischaemia, angiogenesis,
CC nervous system disorders e.g. Alzheimer's disease, infections caused by
CC bacteria, viruses and fungi and ocular disorders e.g. corneal infection,
CC and many other disorders listed in the specification. The polypeptides
CC can also be used to aid wound healing and epithelial cell proliferation,
CC to prevent skin aging due to sunburn, to maintain organs before
CC transplantation, for supporting cell culture of primary tissues, to
CC regenerate tissues and in chemotaxis. The polypeptides can also be used
CC as a food additive or preservative to increase or decrease storage
CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,
CC minerals, cofactors and other nutritional components. The present
CC sequence represents a novel secreted protein of the invention. Note: The
CC sequence data for this patent did not form part of the printed

Query Match 8.9%; Score 712; DB 4; Length 141;
Best Local Similarity 99.3%; Pred. No. 4.1e-41;
Matches 140; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 464 SSKPHKHLPPAALHLIAYKENKREDKRSALSCVISTARLLSSEDRARLPPEELRSIVQ 523
Db 1 SSKPHKHLPPAALHLIAYKENKREDKRSALSCVISTARLLSSEDRARLPPEELRSIVQ 60
QY 524 KRYELLEHKRWASMSSEQRKEYLKKKREELKKKLKEXAKERREKEMLERLEKQKRYEDQ 583
Db 61 KRYELLEHKRWASMSSEQRKEYLKKKREELKKKLKEXAKERREKEMLERLEKQKRYEDQ 120
QY 584 ELTGKNLPAFRLVDPPEGLPN 604
Db 121 ELTGKNLPAFRLVDPPEGLPN 141

RESULT 32
ABU5250
ID ABU5250 standard; protein; 141 AA.
XX
AC ABU5250;
XX
XX 18-MAR-2003 (first entry)
XX
XX Human novel polypeptide #337.

DE Human; neural disorder; immune system disorder; renal disorder;
KW muscular disorder; respiratory disease; reproductive disorder;
KW gastrointestinal disorder; pulmonary disorder; cardiovascular disorder;
KW hyperproliferative disorder; inflammatory disease; allergic reaction;
KW blood related disorder; cancer; immunosuppressive; antiinflammatory;
KW cardiovascular; nephrotropic; cytostatic; antiallergic; thrombolytic;
KW haemostatic; antiarteriosclerotic.
XX
OS Homo sapiens.
XX
XX US2002132753-A1.
XX
XX 19-SEP-2002.
XX
XX 17-JAN-2001; 2001US-00764864.
XX
XX 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 28-JUN-2000; 2000US-0214886P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.

PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 22-AUG-2000; 2000US-0226868P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 08-SEP-2000; 2000US-0231413P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 27-SEP-2000; 2000US-0235834P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 13-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241809P.
PR 01-NOV-2000; 2000US-0244617P.
PR 17-NOV-2000; 2000US-0249299P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.

XX (ROSE/) ROSEN C A.
PA (RUBE/) RUBEN S M.
PA (BARA/) BARASH S C.
XX
PI Rosen CA, Ruben SM, Barash SC;

XX WPI; 2003-147444/14.
XX N-PSDB; ABX73509.

XX New polypeptides and nucleic acids, useful in gene therapy for treating,
PT inhibiting or preventing e.g. neural, immune system, muscular,
PT respiratory, reproductive, gastrointestinal, pulmonary, cardiovascular or
PT renal disorders.

XX Claim 11; SEQ ID NO 1134; 402pp; English.

XX The invention relates to human novel polypeptides and their associated
CC polynucleotides. The polypeptides and polynucleotides are useful in gene
CC therapy for treating, inhibiting or preventing neural disorders, immune
CC system disorders (e.g. systemic lupus erythematosus, rheumatoid arthritis
CC and multiple sclerosis), muscular disorders, respiratory diseases (e.g.
CC nasal vestibulitis, nasal polyps and sinusitis), reproductive disorders,
CC gastrointestinal disorders, pulmonary disorders, cardiovascular disorders
CC (e.g. congenital heart defects, Einstein's anomaly and hypoplastic left
CC heart syndrome), renal disorders (e.g. acute kidney failure and end-stage
CC renal disease), hyperproliferative disorders (e.g. Hodgkin's disease and
CC leukaemia), inflammatory diseases (e.g. septic shock, bursitis and

CC appendicitis), allergic reactions and conditions (e.g. asthma), blood
CC related disorders (e.g. thrombosis, atherosclerosis and myocardial
CC infection) and cancerous diseases. Sequences ABU54914-ABU55699 and
CC ABU55748 represent human novel polypeptides of the invention
XX
SQ Sequence 141 AA;

Query Match 8.9%; Score 712; DB 6; Length 141;
Best Local Similarity 99.3%; Pred. No. 4.1e-41;
Matches 140; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 464 SSKPHKLPAAHLIAYKENKDRKDSALSCVISTARLLSSDRARLPBELSLVQ 523
DB 1 SSKPHKLPAAHLIAYKENKDRKDSALSCVISTARLLSSDRARLPBELSLVQ 60
QY 524 KRYELLEHKRWASMEEQKEYLKKKRELLKKKKEKAKERREKEMLERLEKQRYEDQ 583
DB 61 KRYELLEHKRWASMEEQKEYLKKKRELLKKKKEKAKERREKEMLERLEKQRYEDQ 120
QY 584 ELTGKNLPAPRLVDTPEGLPN 604
DB 121 ELTGKNLPAPRLVDTPEGLPN 141

RESULT 33
ABB58706
ID ABB58706 standard; protein; 1476 AA.
XX
AC ABB58706;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 2910.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
FN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US009231.
XX
PR 23-MAR-2000; 2000US-0191637P.
PR 11-JUL-2000; 2000US-00614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR N-PSDB; ABL02809.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signaling and cell-cell
PT interactions.
XX
PS Disclosure; SEQ ID NO 2910; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABLI6176-ABLI30511), expressed DNA
CC sequences (ABLI01840-ABLI6175) and the encoded proteins (ABB57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 1476 AA;

Query Match 8.4%; Score 668; DB 4; Length 1476;
Best Local Similarity 18.4%; Pred. No. 1.3e-36;
Matches 332; Conservative 255; Mismatches 504; Indels 710; Gaps 56;

QY 3 PLLGRKPPPLVNPPLGCEBPF-----FTIPHTQEAFTREEYEARLERYSERIWTCKSTG 56
DB 2 PICKREGFDLNQEGKNETFHDNDQVFCYITKRIFRDYEHYFRHVWVINSTVWQCEATG 61
QY 57 SSQLTHKEAWEBEQVAELLKEEFPANWEKLVLEWVHNTASLEKLVDTAWLEIMTK-YA 115
DB 62 KENLYTEEAVKSER-AARKKMEQFKQSLRAPVLLVVEHAQSAVNTLNMIVAKFLRKRYF 120
QY 116 VGEEDCDFEYGEKMKLVKIVHIPLEKVDDEATEKSDGACDPSDSENSSOIAQDHOK 175
DB 121 IGEVSVQAKKATYTVLGKL-----DKMPEPLNGIYED 156
QY 176 KETVVKDEGRRESINDRARRSPKLPKLSKKGERKWAAPPKFLPHKYDYKLQNEQKIISN 235
DB 157 TDNLVY-----RLRPNKGDPKSAE-----LD 176
QY 236 VPADSLIRTERPPNKEIVYFPIRHNAIRAGTGENAPWVEDELVKYKSLPSKFSDDLDP 295
DB 177 LPFRQLRRSRMEFNLENLSMFIKSNVSRV-----DGLLR--PKPEAYKQYVTD 223
QY 296 YKYMTLNSTKRKNTGSPDRKPSKSKTDNSSLSPNPKLWCHVHLKSLSGSPKLVKN 355
DB 224 ----GVNFS-----IPGKMPYSPAKIK-- 244
QY 356 SKNKSPEEHLSEMMKMMSPNKLHTNFHIPKGPAPKPGKHSKPLKAKGRSGILNQ 415
DB 245 -----KPDGK-----KQSTLNKYIVAGE 262
QY 416 KSTGSKSPKGLKTPKTKMKTLLDMAKGTQKMTAPRNSGGTPTSSKPHKLPPAA 475
DB 263 ATAASKAKAK----- 273
QY 476 LHLIAYYKENKDRKRSALSCVISTARLLSSDRARLPBELSLVQRYELLEHKRW 535
DB 274 -----SAAKSLAEELERV----- 286
QY 536 ASMSEQRKYLKKREELKKLKEKAKERREKEMLERLEKQ-----KRYEDQELTG-KN 589
DB 287 -----KREKAKLIELEKQAKAEKQALIERVENECNLLQTDLERTDQKV 333
QY 590 LPAFLVDTPEGLPNTLFGDVAMVVEFLSCVGLL--LPDAQYPTITAVSLMALSALSA-DKG 646
DB 334 LPRYQIVTL--LPEHLLGDAPFMMREFMHTYTGLLSGIEVFRQNLSEFVEMTRALTARETA 391
QY 647 GFLYLNRYLVILQTL--LQTLQDEIAEDY-----GELG 679
DB 392 G--PLSDILLVLLGTVDLQKEEEECATYLDRAAQTEQPYWMAQAASHLYAKRHPS 449
QY 680 MKLSBIPLTLSVSELVRLCLRRSD--VQESSEG-----SDTDNKGSAAFEDNEVDQDF 732
DB 450 FKVNELPLDALTLSEVLRHLHLLGSAFVNEKAERWRVMYRNGYSSKEPGLRLHLSHI 509
QY 733 LEKLTSEFFELTSEEKQIILTALCHRLIMTYS-VODHMETRQQMSAELEWLERLVLKEE 791
DB 510 LRILKNHSVYQLKFKDILLIRCLMSQI-MTYSGTINLIEERMEQAKARQDLRALVUGE 568
QY 792 NDKKRAEKQKKEWEAKNKENGKVENGLGKTRDKRKRIVKFEPQVDTEADM--ISAVKSR 850
DB 569 NKRLAAVEINRKLL-----TQMHLEVNGVPEBK 597
QY 851 LLAIQAKKERITQEREMKVKLERQAEERIRKHAASAKAFOEGITAKALVMRRPTIGTD 910
DB 598 REALVEKLKKSIAE-----LHAQSDQOH-KRHEIQMLK-----LHSQLFNFLVYLGND 644
QY 911 RNNRYWLFPSDEVPGFLFEKGWVHDSIDYRNHHCKDHTVS----- 951
DB 645 RCTRYKYVL-ESMPGIFVEHS--PDSLD-----TCLEQPIITNKSQIEIRQQSALPKNRKD 696

```
QY 952 -----GDYDCPRSKANLGNASMTQH---GTATEVAVETTTT----- 988
DB 697 LRVYLLKLYGDDK- KTKKAKHSLNKENOEHLRNGSAEPMVDSPEAPTHPELLMC 755
QY 989 -----KQOQ-NLWFLCDQKELDELLNCLHPQIGRESOL----- 1021
DB 756 SGDRCSCIVHSRNGORQWAYIYKAEIDELIKALNPNGLEVEYELLQELSVLRLSLEQH 815
QY 1022 -----KERLEKRYODIHSIHLARK---PNLGLKSCDGNQELLNP----- 1058
DB 816 AKTCPVDLLSLENETMRKKFMAAMES-ETNRKYGEANFGLPNGTDLNEVMRLHLVDRIIQ 874
QY 1059 -----LRSDLI-----EVATRLQKGLGYVE-----ETSE- 1083
DB 875 FENDIYTGDLGRLLKVKDMKWRSDLLGNYDAQCKLQWPGGKLGDEAGSONESHETHEE 934
QY 1084 -----FEARVISLEKLKD-----FGSCVIAL 1104
DB 935 DDGALLGKYARKPYRDPGMYLAASADTKPLPDSDDDEQHTNAVSIPIAVHNMASALLQV 994
QY 1105 QASVIKKFLQ---GFMA---PKQRRKLOSEDSAKTEEVDEKMWVEAKVASALEKWTA 1159
DB 995 EQAIGKRLKEPYGMKKWDPKQEAELACD-----SRLHQWEVS 1033
QY 1160 IREAOFTFSRMHVLLGMDACIKWDMSAENARCKVCPKGGEDKLTLCDCKNAFHLFCLR 1219
DB 1034 LMESTSPAQVFLHLNLDHCIQWRRSTNKSCKVCRGSDPEKMILLDECNAGTHMFCIK 1093
QY 1220 PALYVPPGEGWCPACQAPATARRNRRGRNYTEESASE-----DSEDESDEEEEEEE 1273
DB 1094 PKLRSVPPGNWCNDCVKSGLSNGONEKDKQATKKRKFIVEEDDEATDEEBEKKD 1153
QY 1274 EEEDEYVAGURLRPRKTIIRGHVS- -IPPAARSG-----RRPGKKPHSTRSQPK----- 1322
DB 1154 DMOTDEDAEHENEKHDEVEDDESVTSTPSSSRVNGRILRRPRTPTSRRLTSKEIEBHA 1213
QY 1323 -----APVDDAEVDDELVLQT----- 1338
DB 1214 QEDVDSGDVDDASITAGEDTIEDSEBKVQCKFYDGGEIKVCQRLFFHLECVHLKR 1273
QY 1339 -----KRSRRROS-----LELQKCEEILHKIVKYSFWSFPREPVTREADSY 1380
DB 1334 NNNNNSSVNNHRRSGRRRTNEHPLNSAALYDLLEQIMKHKAAPFLRPLVTSEVPDY 1393
QY 1381 YDVITHPWFQTVQNKSCGSVRSVQEFITDMKQVFTNAEVYVNCRGSHVLSMWKTEOCL 1440
DB 1394 HQIITKTPMDLAKIKSLNNGAYQLNEELLSD:QLVFRNCDDLNVNEGNEYIDAGCQLERFV 1453
QY 1441 V 1441
DB 1454 I 1454
RESULT 34
ID AAG03625
XX AAG03625 standard; protein; 128 AA.
AC AAG03625;
XX
XX 06-OCT-2000 (first entry)
DE Human secreted protein, SEQ ID NO: 7706.
XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping.
XX Homo sapiens.
XX EP1033401-A2.
PN
```

```
XX 06-SEP-2000.
PD
XX
XX 21-FEB-2000; 2000EP-00200610.
PF
XX
XX 26-FEB-1999; 99US-0122487P.
XX (GRST ) GENSET.
PA
XX
XX Dumas Milne Edwards J, Duclert A, Giordano J;
PI WPI; 2000-500381/45.
XX N-PSDB; AAC03631.
DR
XX
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures.
XX
XX Claim 13; SEQ ID NO 7706; 71pp + Sequence Listing; English.
XX
XX The present sequence is a polypeptide encoded by one of a large number of
CC 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs were
CC prepared from total human RNAs or polyA+ RNAs derived from 30 different
CC tissues. EST sequences usually correspond mainly to the 3' untranslated
CC region (UTR) of the mRNA because they are often obtained from oligo-dT
CC primed cDNA libraries. Such ESTs are not well suited for isolating cDNA
CC sequences derived from the 5' ends of mRNAs and even in those cases where
CC longer cDNA sequences have been obtained, the full 5' UTR is rarely
CC included. 5' ESTs are derived from mRNAs with intact 5' ends and can
CC therefore be used to obtain full length cDNAs and genomic DNAs. 5' ESTs
CC are also used in diagnostic, forensic, gene therapy and chromosome
CC mapping procedures. They are used to obtain upstream regulatory sequences
CC and to design expression and secretion vectors
XX
XX Sequence 128 AA;
SQ
Query Match 8.0%; Score 637; DB 3; Length 128;
Best Local Similarity 99.2%; Pred. No. 5.5e-36;
Matches 127; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 680 MKLSEIPLTHSVSELVRLCLRRSDVQBESESGSDTDNKNDSAAAFEDNEVQDEFLEKLETS 739
DB 1 MKLSEIPLTHSVSELVRLCLRRSDVQBESESGSDTDNKNDSAAAFEDNEVQDEFLEKLETS 60
QY 740 EFFELTSEKLIQITLALCHRIILMTYSVQDHMETROQMSAELWKERLAVLKEENDKKRAEK 799
DB 61 EFFELTSEKLIQITLALCHRIILMTYSVQDHMETROQMSAELWKERLAVLKEENDKKRAEK 120
QY 800 QKRKEMEA 807
DB 121 QKRKEMEA 128
RESULT 35
ABM85416
ID ABM85416 standard; protein; 1586 AA.
XX
XX ABM85416;
XX
XX 18-NOV-2004 (first entry)
XX
XX Mouse protein sequence MCP4389.
XX
XX Cytostatic; carcinoma; lymphoma; cancer; murine.
XX
XX Mus musculus.
XX
XX WO2003073826-A2.
XX
XX 12-SEP-2003.
XX
XX 28-FEB-2003; 2003WO-US006235.
XX
```

PR 01-MAR-2002; 2002US-00087192.
 XX (SAGR-) SAGRES DISCOVERY.
 XX Morris DW;
 XX WPI; 2003-328604/31.
 XX
 XX Recombinant nucleic acid useful for diagnosis and treatment of carcinoma
 XX comprises a nucleotide sequence.
 XX
 XX Claim 5; SEQ ID NO 663; Opp; English.
 XX
 XX The present invention relates to novel DNA and protein sequences which
 XX are associated with carcinomas. The sequences are useful for: (i) for
 XX screening drug candidates; (ii) for screening of bioactive agent capable
 XX of binding to Carcinoma Associated protein (CAP); (iii) for screening of
 XX a bioactive agent capable of modulating the activity of CAP; (iv) for
 XX evaluating the effect of a candidate carcinoma drug; (v) for diagnosing
 XX carcinoma; (vi) for inhibiting the activity of CAP; (vii) for treating
 XX carcinoma; (viii) for neutralizing the effect of CAP; (ix) as a biochip;
 XX (x) for diagnosing carcinoma or a propensity to carcinoma; and (xi) for
 XX determining Carcinoma Associated (CA) gene copy number. In addition, the
 XX CA genes are useful as DNA vaccines and the CAP are useful as markers of
 XX carcinoma including lymphoma. The present sequence is one such CAP. Note:
 XX This patent is an equivalent to basic patent US2002182586A1, for which no
 XX sequence data was published
 XX
 XX Sequence 1586 AA;
 SQ
 Query Match 7.5%; Score 599; DB 7; Length 1586;
 Best Local Similarity 20.9%; Pred. No. 8.2e-32;
 Matches 318; Conservative 231; Mismatches 549; Indels 422; Gaps 61;
 QY 137 IHPLEKVDDEATE---KSDGACDSSDSKE-----NSSIAQDHQKKEIVKKE 182
 DB IIIHAAEELTSSVVAENGTLVGLSLEEEQPELXMGVNGSVSSVESLHQEVSVLVLPD 295
 QY 183 -----DEGRRESINDRARRSPKLTSLKKGKRWAPPKFLPHKY---DVKLON-EDKI 232
 DB 296 PTVSCIDD--PSHLPQLEDTPILSEDSLEPPDSLAARPEVSGSLYGIDDAELMGAEKDL 353
 QY 233 -----ISNVPADSLIRTERPPNKEIVRYFIRHNALRAGTGGENAPWVDE 277
 DB 354 PLEGNPVISALDCPALSNANAPSLADDQSASIF-----VSPTSPPVLGES 401
 QY 278 LVKYSLSLPSKFSDFLLDPKYMTLNFSTKRNKGTSPDRPKSKSKTDNSLSLPLNPKLW 337
 DB 402 VLOGSPSPSPAAAF-----QTVSPA--RKNVSSAPKARADREETGGAVA----- 444
 QY 338 CHVHLKXSLSGSPKLVKNSKNSKSPHEEHEEMKMSPNKLHTNPHIPK----- 386
 DB 445 -----VSGSDVULK-----RRATPEE-----VRLPLQHWRRREVRIKGSRWQGETWY 489
 QY 387 KGPPAKPGKHSKPKLAKGRSKGIILNGOKSTGNSKSPKGL-----KTPK-TQMKQM 438
 DB 490 YGPCGKRMKQF---PEVIKYSLNRVVHSVRREHFSFSPMPVGDFFPEERDTPEGLQWQL 546
 QY 439 TLLDMAKGTQKMTAPRNSGTTPTSSK-PHKILP-----PAALHLIAYKENKORE 489
 DB 547 SAEIEFSRIQAIT---GKRGRPRNNEKAKNEVPKVRGRGPRPKIKMPELLINKTDNRL 602
 QY 490 DKESALSCVISTARTLLSSDRALPEELSLVQRYELLEHKRWASMSSEORKEYLAK 549
 DB 603 PKK-----LETQETILSEDDKAMTKNKKMKQK--VORGESQTPVQGGARNKRDOTKS 654
 QY 550 -KREELKKLKEKAKE--RREKEMLERLEKQRYEDQELTG-KNLPAFLVDTPEG--LPN 604
 DB 655 LKQKDTKKLKKLESEQRQQAILE--EMKKTPEDCMLSDHQPLPDTFRI---PGLTSS 709
 QY 605 TLFQGVAMVVEFLSCYSGLLLPDAQYPTITAVS-LMEAL-----SADKGGFLYNRLVIL 658
 DB 710 RAPSDCLTIVEFLHSFGKVLGFDLTAKDVPISGLVQLQGLLQCGDSLDK-----VQDL 760

RESULT 36

AAW81171

ID AAW81171 standard; protein; 1972 AA.

XX AAW81171;

XX

DT 05-MAY-1999 (first entry)

XX DE Human BAZ2-beta protein.
 XX KW Transcriptional regulator; transcription; BAZ1-alpha; bromodomain; BAZ;
 KW atypical zinc finger; testis; human; tumour; BAZ1-beta; BAZ2-alpha; drug;
 KW BAZ2-beta; treatment; cancer; proliferative disorder; screening.
 XX OS Homo sapiens.
 XX PN WO9847920-A1.
 XX PD 29-OCT-1998.
 XX PF 17-APR-1998; 98WO-JP001783.
 XX PR 18-APR-1997; 97JP-00116570.
 XX PR 24-OCT-1997; 97JP-00310027.
 XX (CHUS) CHUGAI RES INST MOLECULAR MEDICINE INC.
 XX PI Jones MH;
 XX DR WPI; 1998-583603/49.
 XX DR N-PSDB; AAV68403.
 XX CC Transcriptional regulator gene family containing bromodomain - may be
 PT expressed in testis tissue and is useful for treatment of cancer and
 PT other proliferative disorders.
 XX PS Claim 1; Page 100-116; 187pp; Japanese.
 XX CC This sequence represents the human BAZ2-beta protein, a member of a
 CC family of transcriptional regulator genes containing a bromodomain (BAZ,
 CC Bromodomain with Atypical Zinc finger) which are expressed specifically
 CC in testis tissue and also in certain tumour lines. Transgenic cells may
 CC be used for the preparation of the BAZ1-alpha, BAZ1-beta, BAZ2-alpha and
 CC BAZ2-beta proteins. These proteins can be used in the treatment of cancer
 CC and other proliferative disorders, and in screening of compounds for
 CC their binding ability to the expression products (e.g. for use as drugs
 CC by modulation of transcriptional regulation)
 XX SQ Sequence 1972 AA;
 Query Match 6.7%; Score 533; DB 2; Length 1972;
 Best Local Similarity 18.0%; Pred. No. 4.1e-27;
 Matches 353; Conservative 256; Mismatches 598; Indels 758; Gaps 68;
 QY 8 KPFPVLPVGPPEF-FTIPH-----TOAFRTREYEYEARLERYSERIWTCKST 55
 DB 189 KPLSLVNQAKETVMKLIIVPSPDVLKAGNKNTSESSLTSELRSKREY-----KQA 241
 QY 56 GSSQLTHKAWBEEQEVALLK-----BEFPAYVKLVLEVMVHN-----TA 97
 DB 242 FPSQLKQBSKSLKKVIAALSNPRKATSSSPA-HPKQTLNHNHPFLTNALLGNHPNG 300
 QY 98 SLEKLVDTAWLIMTKYAVGECDEP-----124
 DB 301 VIQSVIQEAPLALTTRKQSKINENIAAASPTPSPVNLSTSGRTPGNTPVWPSAS 360
 QY 125 -----GKEKML--KVKIVKI-----HPL-----EKVDEE 146
 DB 361 PILHSQGEKAVNNVNPVKTOHSHPAKSLVEQPRGTSDIPSSKDSNDEEEDDE 420
 QY 147 ATEKSDGACDSPSKENSQIAQHQKKTIVKEDEGRRESINDRARRSPKRLPTSLK 206
 DB 421 EDEDEDEDDSDQSDSNSDSTEGSEEDDDKQDSDS-----T 467
 QY 207 KGERKWPAPKPLPHKYDVKLQNEDKIISNV--PADSLITERPPNKEIVRYFIRHNALRA 264
 DB 468 EGE-----KTSMKL-----NKTTSVKSPSLSLGHSTFRNLHTAK-----503
 QY 265 GTGENAPWVVEDELVKYSLPSKFSDFLLDPKYMTLPNSTKRKNTGSPDRKPSKSK-T 323

DB 504 -----APGSAPAALCSSESQSPA-----FLGTSSSTL-----TSSPHSGTSKRRVVT 544
 QY 324 DNSLSLSPNLKLMCHVHLKKSLSGSLPKVKNKSNKSP-----BEHLEEMMMKMSPN--- 376
 DB 545 DERELRIPLEYG-WQRETRIRNFGG---RLQGEVAYYAPCGKKLRQYPEVIKYLSRNGIM 600
 QY 377 -----KLHT-NPHIPKGP-----PAKKPG 395
 DB 601 DISRDNFSFSAKIRVGDYFYEARDGPFQEMQWCLLKEDVI PRIRAMEGRRGPPPNPDRQA 660
 QY 396 KHSKPLKAKORSKI-----LNGQKSTGNS 421
 DB 661 REESRMRRKGRPPNVNNAEFLDNADAKLKLQAOETARQAQIKLRLKQKQARVA 720
 QY 422 KSPKK--GLKTPKTKMQMTLLDMAKGTOKMTRAPRNSGGTPRTSSSKPHKLPALHLI 479
 DB 721 KEAKKQQAIAAEERKQKQKQIKMKQOEKIRIQ-----IRNEKEL--RAQIIL 769
 QY 480 AYYKENKDREDKRSALS CVISKTARLLSSEDRARLPEELRSLVQKRYELLEHHKRWASMS 539
 DB 770 EAKKKKEE-----AANAKLLEAEKRIK-EKEMR--RQOAVLLAKHOER-----809
 QY 540 EEOQKEYLKKREELKKLKKAKERREKEMLEKRYEDQEL---TGKNL-----590
 DB 810 ERRQHMMLKAMEARKKAEKERLKQEKDEKLNKERKLEQRRLEMAKELKKPNED 869
 QY 591 -----PAFLVDTPB--GLPNTLFGDVAMVVFSLSCYSGLLLPDPAQYPI TAVLSMEALS 642
 DB 870 MCLADQKPLPELPRIPLGLVLSGSTFSCLMVQVFLRNFGKVLGPDVNDVNLVQLB-- 927
 QY 643 ADKGGFL-----YLNRLVILLQTLLOLLOEIAEDY---GELGMKLSIEPLTLHSVSEL 695
 DB 928 ----GLLNIGDSMGVQDILLVRLLSAAVCDPGLITGYKAKTALGHLNLVGNRDNVSEI 983
 QY 696 VRLCLRRSDVOESEGSDTDNDKDSAPEDNEVDQEFLEKLETSBFFLTSSEKLIQILTA 755
 DB 984 LQIFW-----EAHCGOTELTESLTKTAPQAHHTPAQKASVLAF 1020
 QY 756 LCHRILMTYSVDHMETPQMSAEIWKERLAVLKEENDKKAERKQKREMAKKNKGV 815
 DB 1021 LINEIACSVSVSEIDKNIDYMSNLRDKWV---EGKLRKLIHAKTKGRTSGI 1076
 QY 816 ENG-----LG-----KTDRKRIVKFPFQVDTEADMSIAVKRRLLAIQAKEREIOE 864
 DB 1077 DLGEEQHPGTPGCKERRKGGSDYDDDDDDSD-----QGEDDEDEE 1123
 QY 865 --REMKVKLEQABEERIRKHAEEK-----AFQEGIAKAKLVWRTPITGT 909
 DB 1124 DKEDQKGGKTIDCEDEDEGQAASVEELEKQIEKLSKQSQYRRKLFDAHSLRSVMFGP 1183
 QY 910 DRNHNRYWLPFSDVEVPGLFIE-----KGWVH-----DSIDY 939
 DB 1184 DRYRRRYWIL--PROCGIFVEGWESGEGLEETAKREKKAESVOIKEEMFETSGDSLNC 1242
 QY 940 RFNHHC-----KDHT---VSGDEDYCPRSK-----KANLGKNA---969
 DB 1243 SNTDCEQEKEDLKEKDNLTNLFQKPGSPKLSKLELVAKMPPESEWMTPKNAGANGCTL 1302
 QY 970 -----SMNTQHGTA TEAVE-----TTTTPKQGG 992
 DB 1303 SYVNSGKHSLSGVQSTATQSNVKEADNNLNTGSSGPGKFPYPLPNDQLLTKLRNQ 1362
 QY 993 NLWFL-----CD-----999
 DB 1363 --WFSLRPTPCDDTSLTHADMSTASLVTPOSPPSKSPSPAPLGSSAQNPVGLNPPA 1420
 QY 1000 -----999
 DB 1421 LSPLOVKGVSNMGLQFCGWPTGVVTSNI PFTLSVPSLGSGLSGLSEGNNSFLTSNVASS 1480
 QY 1000 -----SOKELDELNLCHLP 1013
 DB 1481 KSESPVQNEKATSAQPAAEVAKVDVDFSPKPIPEEMQFGWIRIIDPEDLKALLKVLHL 1540

Db 721 KEAKQQAIAAEKKEKQEIKIMKQEKIKRIQQ-----IRMEKEL--RAQQIL 769
QY 480 AYYKENKORERKRSALSCVISKTAKLSSSDRRLPEELRSUVKRYELLEHHKRWASMS 539
Db 770 EAKKKKEE-----AANAKLLEAEKRIK-EKEMR---RQAVLLKHQER----- 809
QY 540 EQORKEYLKKKEELKKLKEKAKERREKEMLERLEKQKRYEDQEL---TGKVL----- 590
Db 810 ERRQOHMLMKAEARKAEERKLEKQEKRDKRLNKERKLEQRLELEMAKELKPNED 869
QY 591 -----PAFRVLVDTPK-GLPNTLPGDVAMVVFSLSCYSGLLLPDAQYPIITAVSLMEALS 642
Db 870 MCLADQKPLPELPRIFGLVLSGSTFSDCLMVVQFLRNFQKVLGDFVDVFNLSVLQE--- 927
QY 643 ADKGFL-----YLNRLVILLQTLLOTLQDIBADY---GELGKMLSEIPIULHSVSEL 695
Db 928 -----GLLNIGDSMGVEQDILLVRLLSAAVCDPLITGYAKTALGEHLLNVGVNRDNVSEI 983
QY 696 VPLCLRRSDVQSESGSDTDNDKDSAAPEADNEVDQEFLEKLETSBFFELTSEKQLILTA 755
Db 984 LQIFW-----EAGCGQTELTSKTKPAQHTPAQKASVLAP 1020
QY 756 LCHRIILMTYSVDHMETROQMAELWKERLAVLKEENDKRAEKQKREMAKNGKGV 815
Db 1021 LINELACSKSVSEIDKNDIDYMSNLRDKWV---EGKLRKLRILHAKTKGRTSGI 1076
QY 816 ENG-----LG-----KTRKRIIVKFPQVDTEADNISAVKSRRLLAIOAKKERETOE 864
Db 1077 DLGEBQHPLGTPTPKRRKRGKGGSDYDDDDDDSD-----QGEDDEDEE 1123
QY 865 ---REMKVKLERQAEERIRKHAAEK-----AFQEGIAKAKLVRRTPGT 909
Db 1124 DKEDQKGRKTDICEDEDEGQAASVEELEKQIEKLSKOQSYRRKLFASHSLRSVWFGP 1183
QY 910 DRNHNRYLFSDEVPLFTIE-----KGWVH-----DSIDY 939
Db 1184 DYRRRYWIL-PRCGGIFVEGMESEGLEEIAKERELKKAESVQIKEMFPTSGLSNC 1242
QY 940 RFNHHC-----KDHT---VSGDEDYCPRSK-----KANLGRNA--- 969
Db 1243 SNTDCEQKEDLKEKDNLTFLQKPGSFSKLSKLEVAKMPPSEVMTPKPNAGANGCTL 1302
QY 970 -----SMNTQHGTAETAVE-----TTTPKQOG 992
Db 1303 SYONSGKHSGLSVQSTATQSNVEKADSNLFTNGSGPGKFSYPLNDQLLTKLTKNRQ 1362
QY 993 NLWFL-----CD----- 999
Db 1363 --WFSLLPTPCDDTSLTHADMSTASLVTPQSPSKSPPTPAPLGSSAQNPVGLNPPA 1420
QY 1000 ----- 999
Db 1421 LSPLOVKGVSMWGLQFCGWPTGVVTSNIPFTLSVPSLGLSLGSEGNGNSPLTSNVASS 1480
QY 1000 -----SQKELDELINCLHP 1013
Db 1481 KSESPVQNEKATSAQPAAVEVAKVPDPPSPKPIPEEMQFGWRIIDPEDKALLKVLHL 1540
QY 1014 QGIRSQLERLEKRYQDIHSHILARKPNGLKSCDGNQELLNFRSLDLIE---VATRL 1070
Db 1541 RGIREKALQKQIQK-HLDYITQACLKXK-DVAIIELNENE--NQVTRDIVENWSVESQA 1596
QY 1071 QKGGLGYVEETSEFARVVISLEKLKDFGECVIALQASVIKFLQGFMAPKQ----- 1121
Db 1597 MEMDLSVLQOVEDLERRVAS-----ASLQ-----VKGMWCPPEPASEREDLVY 1638
QY 1122 -----KRRKLQSDSAKTEVDEEKKQWE--EAKVAS 1151
Db 1639 FEHKFTKLCBHDGEFTGEDESSAHAEURKSDNPLDIAVTRLADLERNIERRIEDIAP 1698
QY 1152 ALEKWKTAIREAQTSRHHVLGLMGLDACIKWDSAEANARCKVCPKKGDDKLILDCGNK 1211

Db 1699 GLRVRRRALSEARGAAQVALCIQLOKSIWEKSIKMYVQICRKGDNELLLLLCDGCDK 1758
QY 1212 AFHLFCLRPALYEPDGEWQCPACOPATA-----RRNSRGRNTEESASE-----DS 1258
Db 1759 GCHTYCHRPKITTPDGDWFCPACIAKASGOTLKIKKLHVKGKKTNESKKGKKVTLTGDT 1818
QY 1259 EDDSDDEEBEEEEEDVEVAGRLRPRKTIRGKHSVIPPAARSRRPGKPKPHSTR 1318
Db 1819 EDEDS-----ASTSSSLKRGKDLQKRX 1841
QY 1319 SQKAPPVDDAEVDLVLQTKRSRRRQLEKQCEILHLKIVKYRFSWPPFPVTRDEAE 1378
Db 1842 MEENTS--INLSKQESFTSVKPKRDDSCKDLALCSMLTETMETHEDAMPFLLPVNLKLP 1899
QY 1379 DYVDVITHPMPQFOTVONKSCGSYSRVSQVFLTDMKQVFTNAEVN 1423
Db 1900 GYKVIKPKMPDFSTIREKLSGQVPLNLETFALDVLRLVDFDNCETEN 1944
RESULT 38
ADP54420
ID ADP54420 standard; protein; 1972 AA.
XX
AC ADP54420;
DT 18-NOV-2004 (first entry)
XX
DE Human PRO protein sequence SEQ ID NO:396.
KW human; PRO; immune related disease; inflammatory immune response;
KW immune response stimulation; antiallergic; antianemic; antiarthritic;
KW antiasthmatic; antidiabetic; antiinflammatory; antiapocretic;
KW antirheumatic; antichyroid; CNS; dermatological; gastrointestinal;
KW haemostatic; hepatotropic; immunostimulant; immunosuppressive; muscular;
KW nephrotropic; neuroprotective; osteopathic; respiratory; vasotropic;
KW virucide; gene therapy.
XX
OS Homo sapiens.
XX
FN WO200403956-A2.
XX
PD 13-MAY-2004.
XX
PF 28-OCT-2003; 2003WO-US034381.
XX
PR 29-OCT-2002; 2002US-0422472P.
XX
PA (GETH) GENENTECH INC.
XX
PI Aggarwal S, Clark H, Gurney AL, Schoenfeld J, Williams PM;
PI Wood WI, Wu TD;
XX
DR WPI; 2004-376182/35.
XX
DR N-PSDB; ADP54419.
XX
PT New PRO polynucleotides and polypeptides, useful in useful in diagnosing
PT and treating an immune related disease, e.g. systemic lupus
PT erythematous, rheumatoid arthritis, diabetes mellitus or asthma and in
PT stimulating an immune response.
XX
PS Claim 1; SEQ ID NO 396; 3009pp; English.
XX
CC The present invention describes an isolated PRO nucleic acid (I). Also
CC described: (1) a vector comprising (1); (2) a host cell comprising the
CC vector of (1); (3) a process for producing a PRO polypeptide; (4) an
CC isolated PRO polypeptide; (5) a chimeric molecule comprising the
CC polypeptide of (4) fused to a heterologous amino acid sequence; (6) an
CC antibody which specifically binds to a polypeptide of (4); (7) a
CC composition of matter comprising a polypeptide of (4), an agonist or
CC antagonist of the polypeptide or an antibody that binds to the
CC polypeptide in combination with a carrier; (8) an article of manufacture
CC comprising a container, a label on the container and a composition of
CC matter of (7); (9) a method of treating an immune related disease in a

CC mammal; (10) a method for determining the presence of a PRO polypeptide
 CC in a sample suspected of having the polypeptide; (11) a method of
 CC diagnosing an immune related disease or an inflammatory immune response
 CC in mammal; (12) a method of identifying a compound that inhibits or
 CC mimics the activity of or expression of a gene encoding a PRO polypeptide
 CC ; and (13) a method of stimulating the immune response in a mammal. The
 CC PRO sequences have anti-allergic, anti-inflammatory, anti-arthritis,
 CC anti-asthmatic, anti-diabetic, anti-inflammatory, anti-psoriasis,
 CC anti-rheumatic, anti-thyroid, CNS, dermatological, gastrointestinal,
 CC haemostatic, hepatotropic, immunostimulant, immunosuppressive, muscular,
 CC nephrotropic, neuroprotective, osteopathic, respiratory, vasotropic and
 CC virucide activities, and can be used in gene therapy. The nucleic acid
 CC (I) and the encoded polypeptides, compositions, kits and methods are
 CC useful in diagnosing and treating an immune related disease and in
 CC stimulating an immune response. The present sequence represents a human
 CC PRO protein from the present invention.
 XX
 SQ Sequence 1972 AA;

Query Match 6.7%; Score 533; DB 8; Length 1972;
 Best Local Similarity 18.0%; Pred. No. 4.1e-27;
 Matches 353; Conservative 256; Mismatches 598; Indels 758; Gaps 68;

QY 8 KPFLVNPPLGPEPF-FTIPH-----TQAFRTREYERLERYSIWTCST 55
 DB 189 KPFLSLVNOAKKTYMKLVIPSPDLKAGNKNTSESSLTSLRSKREY-----KQA 241

QY 56 GSSQLTHKEAWEEOEVBELLK-----EFPAYWEKLVLEVMHNN-----TA 97
 DB 242 FPSQLKQESSKSLKVIAALSNPKATSSPA-HPKQTLNHNHPFLTNALLGNHPNG 300

QY 98 SLEKLVDTAWLIMTKYAVGECDFEV-----S-----124
 DB 301 VIQSVIQEAPLALTTKMQSKINENIAAASSTPSSPVNLSTSGRRTPGNQTPWPSAS 360

QY 125 -----GKEKML--KVIKVI--HPL-----EKVDEE 146
 DB 361 PILHSQGEKAVNNVNPVKTHGHPAKSLVEQFRGTDSDIPSSKSDSDNEDEE 420

QY 147 ATEKSDGACDPSDKENSSQIAODHOKETVWEDEGRRESINDRARRSPKLPSTSLK 206
 DB 421 EDEDEDEDESDSQSDSNSDSTEGSEEDDDDDQDESDD-----T 467

QY 207 KGERKWAPPFLPHKYDVKLQNEKIIISNV--PADSLINTERPPNKEIVRYFIRNALRA 264
 DB 468 EGE-----KTSMKL-----NKTSSVKSPSMLTGHSTPRNLHIAK-----503

QY 265 GTGENAPVWVEDELVKKYSLSKFSDFLLDPKYMTLNPSTKRKNTGSPDRKPSKSK-T 323
 DB 504 -----APGSAPALCESQSPA-----FLGTSSTL-----TSSPHSGTSKRRRTV 544

QY 324 DNSLSLSPNLKLVHVKLSGLSPKLVKNSKNSKP-----EEHLEEMKMSNP-----376
 DB 545 DERELRIPLEYG-WQRETRIRNFGG--KLQGEVAYAPCGKKLRQYPEVITYLSRNGIM 600

QY 377 -----KLHT-NFHIPKPGP-----PAKKPG 395
 DB 601 DISRDNFSFSAKIRVGDVFEARDGQEQMWCLLKEEDVIPRIAMEGRGRPPNDRQRA 660

QY 396 KHSIDPLKAKGRSGKI-----LNQKSTGNS 421
 DB 661 REESRMRRKGRPPNVGNAEFLDNADAKLLRLKLOAQETARQAQIKLLRLKQKQBARVA 720

QY 422 KSPKK--GLKTPKTKMQWTLDMAGTQKMTAPRNSGGTPTRTSSKPKHLPPAALHLI 479
 DB 721 KEAKQQAQIMAAEERKQEQIKIMKQEQEKIRIQO-----IRMEKEL--RAQQIL 769

QY 480 AYYKENKDREDKRSALSCKVISKTALLSSEDRARLPEELSLVOKRYELLEHKKWASMS 539
 DB 770 EAKKKKEE-----AANAKLLEAKKIK-EKEMR-----ROQAVLLKHQER-----809

QY 540 EQQRKEYLKKRBEELKKLKEKAKERRERKEMLERLEKQRYBDQEL-----TGKNL-----590

DB 810 ERRROHMLMKAMEARKKAABEKERLQEKRDDEKRLNKERLQORLEMAKELKPKNED 869
 QY 591 -----PAFRLVDTPE--GLPNTLFGDVAMVVEFLSCYSGLLPDAQYDITAVSLMEALS 642
 DB 870 MCLADQKPLPELPRIPLGLVLSGSTSDCLMVVQFLRNFGKLVDFVNDVPLSVLQOE-- 927
 QY 643 ADKGQFL-----YLNRVVLITLQTLLODIABDY---GELGMKLSIPITLHSHVSEL 695
 DB 928 -----GLLNTGDSMGVQDILLVRLLSAAVCDPLITGYKAKTALGEHLLVNGVNRDNVSEI 983
 QY 696 VLLCLRRSDVQSESESGSDDDNDKNSAEPEDNEVDQEFLEKLETSFFELTSEKLOILTA 755
 DB 984 LQIFM-----EAGCGQTELTSLTKKPAQHAHTPAQKASVLAP 1020

QY 756 LCHRILMTYSVDHMETRQMSAEIWKERLAVLKEENDKKRAEKQKREKEMAKNKENGKV 815
 DB 1021 LINELACSKSVVSEIDKNIDYMSNLRRDKVV-----EGKLRKLRIRIHAHKTCKRDTSGGI 1076

QY 816 ENG-----LG-----KTDRKRIIVKFPQVDTEADMISAVKSRRLLAIAQAKEREIOE 864
 DB 1077 DLGEBQHPLGTPTPGRKRRKGGSDYDDDDDDDDDD-----QGDDEDEDE 1123

QY 865 --REMKVKLERQAEERIRKHKAAAEK-----AFQEGTAKAKLVNRRTPIGT 909
 DB 1124 DKEDQKGGKTIDCEDEDEGDAQASVEELEKQIEKLSKQOSQYRRKLFDAHSLRSVMFPG 1183

QY 910 DRNHNRYWLFSDDEVPLGFIE-----KGWVH-----DSIDY 939
 DB 1184 DRYRRRYWIL--PRCGIIFVEGMESEGLEETAKEREKUKAESVQIKEMFETSGDSLNC 1242

QY 940 RFNHHC-----KDHT---VSGDEDECPRSK-----KANLGKNA-----969
 DB 1243 SNTDCEQEKEDLKEKONTNLFLQPGSFKLSKLEVAKMPPESEVMTPKPNAGANGCTL 1302

QY 970 -----SMATQHGATATEAVE-----TTTTPKQCO 992
 DB 1303 SYQNSGKHSLSGVQSTATQSNVEKADSNLFWTSGSGFKFYSLPNDOLLTLTEKNRQ 1362

QY 993 NLWFL-----CD-----999
 DB 1363 --WFSLLPRTPCDDTSLTHADMSTASLVTPOSPKSPSPTPAPLGSSAQNPVGLNPPA 1420

QY 1000 -----999
 DB 1421 LSLPQVKGVSVMGLQFCGWPTGVVTSNIPFTLSVPSLGLSGLSEGNCSFLTSNVASS 1480

QY 1000 -----SOKELDELLNCLHP 1013
 DB 1481 KSESPVQNEKATSAQAAAVEYAKVPDFPSPKPIPEEMQFGWRIIDPEDLKALLKVLHL 1540

QY 1014 OGIRESQLKERLEKRYQDI IHSIHLARKPNLGLKSCDGNOLNLFRLSDLE--VATRL 1070
 DB 1541 RGIREFALQKQIQK--HLDYITQACLNK--DVAITELNENE--NQVTRDIVENWSVEBOA 1596

QY 1071 QKGGLGYVEETSEFEARVISLEKLDKDFGECVIALQASVIKKFLOGFMAPKQ-----1121
 DB 1597 MEMDLSVLQVEDLERVAS-----ASLQ-----VKGMCMPEPASEREDLVY 1638

QY 1122 -----KRRKLQSEDSAKTEEVDEEKKQWVE--BAKVAS 1151
 DB 1639 FEHKSFTLCKEHDGEFTGEDESSAHALERKSDNPLDIATVRLADLERNIERREDIAP 1698

QY 1152 ALEKWKTAIRAQATFSRMHVLGLMDACIKWDMASNAENARCKVCPKKGDDDLILCDECNK 1211
 DB 1699 GLRVRRALSARSAAQVALCICQLOKSIAMEKSIIMKYVCQICRKGDNELLLCGGCDK 1758

QY 1212 AFHLFCLRPALYVDPDGEWQCPACOPATA-----RNSRGRNYTESAS--DS 1258
 DB 1759 GCHTYCHRPKITTPDGDWFCFACITAKASGQTLKIKLHVKGKKNESKGGKVTLTGDT 1818

QY 1259 EDDSDSEEBEEBEEBEEDYEVAGLRURPRKTRIRKSHSVIPPAARSRRRPPQKPHSTRR 1318
 DB 1819 EDEDS-----ASTSSSLKRGKNDLQKGG 1841

QY 1319 SOPKAPPVDDAEVDELVLQTKRSSRQSLQKCEIILHKIVKYRFSWPPREPVTDRDEA 1378
Db 1842 MEENTS--INLSKQBSFTSVKPKRDDSKDLALCSMLITETHEDAWPFLPVLNKLVP 1899
QY 1379 DYYDITHPMDFOITVQNKSCGYSRSVQEFLLTDMKQVFTNAEVYN 1423
Db 1900 GYKVIKKPMDFTSIREKLSGQYFNLETFALDVLVDFNCEFTFN 1944

RESULT 39
ABM85417
ID ABM85417 standard; protein; 1873 AA.
XX
XX ABM85417;
AC
XX 18-NOV-2004 (first entry)
XX
XX Human protein sequence hCP42010.
DE
XX Cytostatic; carcinoma; lymphoma; cancer; human.
KW
XX Homo sapiens.
OS
XX WO2003073826-A2.
PN
XX
XX 12-SEP-2003.
PD
XX
XX 28-FEB-2003; 2003WO-US006235.
PF
XX
XX 01-MAR-2002; 2002US-00087192.
PR
XX
XX (SAGR-) SAGRES DISCOVERY.
PA
XX Morris DW;
PI
XX WPI; 2003-328604/31.
XX
XX Recombinant nucleic acid useful for diagnosis and treatment of carcinoma
PT comprises a nucleotide sequence.
XX
XX Claim 5; SEQ ID NO 666; Opp; English.
XX
XX The present invention relates to novel DNA and protein sequences which
CC are associated with carcinomas. The sequences are useful for: (i) for
CC screening drug candidates; (ii) for screening of bioactive agent capable
CC of binding to Carcinoma Associated Protein (CAP); (iii) for screening of
CC a bioactive agent capable of modulating the activity of CAP; (iv) for
CC evaluating the effect of a candidate carcinoma drug; (v) for diagnosing
CC carcinoma; (vi) for inhibiting the activity of CAP; (vii) for treating
CC carcinoma; (viii) for neutralizing the effect of CAP; (ix) as a biochip;
CC (x) for diagnosing carcinoma or a propensity to carcinoma; and (xi) for
CC determining Carcinoma Associated (CA) gene copy number. In addition, the
CC CA genes are useful as DNA vaccines and the CAP are useful as markers of
CC carcinoma including lymphoma. The present sequence is one such CAP. Note:
CC This patent is an equivalent to basic patent US2002182586A1, for which no
CC sequence data was published
XX
SQ Sequence 1873 AA;

Query Match 6.7%; Score 531.5; DB 7; Length 1873;
Best Local Similarity 19.1%; Pred. No. 4.8e-27;
Matches 261; Conservative 183; Mismatches 438; Indels 483; Gaps 40;

QY 415 OKSTGNSKSPKGLTPTKMKOMTLLDMAKGTQ---KOTRAPRNSGGTPTKTSKPHKHL 471
Db 616 QAITCKRGPR---NTEAKYKEVPKVRGRGPPKVKITEL-----LNKTDNRPLKXL 666
QY 472 PPAALHLIAYYKENDKEDKSALSVCISKTKAR-----LLSSEDRALPEEL 518
Db 667 EA-----QETLNEEDKAK-----IAKSKKMKRQKVQRGECCTTQGGARNRKQET 712
QY 519 RSLVQKRYELLEHKKWASMBEQRKEVILKKKREELKKLKE---KAK----- 563

Db 713 KSLQKEAKKSKKEKGKTKQELKEKVKREKKEKVKMEKEEVTAKAPACKADKTLATQ 772
QY 564 ---ERRKEMLERLEKQKRYEDQELTG-KNLPAFLVDTPEG--LPNTLFGDVAMVVEFL 617
Db 773 RRLERQROQMILEEMKKPTEDMCLTDHPLPDFSRV---PGLTLPFGAFSDCLTIFBL 829
QY 618 SCYSGLLLPDAQYPIITAVS-LMEALSADKGGFLYLNRLVILLQTLQLQLQDEIAEDYG 676
Db 830 HSFGKVLGDFPAKDVPSLGVLOEGLLCOGSLGEVQDLLVRLLLKAALHDPFPSPCQSLK 889
QY 677 ELGKMLSEIPLTHSVSELVRLCLRRSDVOBESGSDTDNDKDSAAAFEDNEVDQEFBLK 736
Db 890 ILGEKVSEIPLTRDNVSEILACFLMAYG-----VEPALCDRL 926
QY 737 ETSFFELTSEBKQILTALCHRI-----LMTYSVQDEHMETROQMSAEIWL--KERLAVLKE 790
Db 927 RTQPFQAPQOQKAAVLAFVLHVLNGSTLIINEIDKTLSESMSSYRKKNKWIVEGRLLRK- 985
QY 791 ENDKKRAEKQKKEWEAKNKENGKVENGLGKTRKRVKFEPOVDTEADMISAVKRR 850
Db 986 ---TVLAKRTGRSEVEMEGP-----EECLGRRRSRIMEETS GMEEEESIAA VPGRR 1037
QY 851 LLAIQAKKERIEQBERMKV-KLERQABEERIRKHAARAKAFQEGIAKAKLVMRTPIGT 909
Db 1038 -----GRDGEVDATASSIPELERQIEKLSKR-----QLFFRKLLHSSQMLRAVSLGQ 1086
QY 910 DRNNRYWLFSDEVPLGFIE-----KGWVHDSIDYRF----- 941
Db 1087 DRYRRRYWVL--PYLAGIFVEGTGNLVPBEEVIKKTSLKVAHAASLNPALFMSKMWELAG 1145
QY 942 -----NHCKDHTVSGDEDYCPRSK----- 961
Db 1146 SNTTASSPARARGPRKTKPGSMQPRHLKS--PVRGQDSEQOQAQLPPEAQLHAPQPOQ 1204
QY 962 -----KANLGKN-----ASMNTQHGTTATVAVETTTPKQG- 991
Db 1205 LQLQLQSHKGFLEQEGSPLSLGQSHDLQSQAFSLWSLQTSQSHSLSSSVLTPDSSPGK 1264
QY 992 -----QNLWFLCDSQ----- 1001
Db 1265 LDPAPSQPPEPPEDEAESPDQALWFNISAQMPCNAAPTTPPPAVSDEDQTPSPQOLAS 1324
QY 1002 ----- 1001
Db 1325 SKPMNRPSAANPCSPVQFSPTPLAGLAPKRRAGDPGEMPQSPGLGQPKRRGRPSKFFK 1384
QY 1002 -----KELDELLNCLHPQGIRESOLKERLEKRYQDI 1032
Db 1395 QMEQRYLTQLTQAQVPPPEMCSGWWIWRDPEMLDAMLKALHPRGIREKALHKLHNK-HRDF 1443
QY 1033 IHSIHL--ARKPNLGLKSCDGNQEB-LNLFRLSDLIEVATRLQKGLGVVEETSEFEARVI 1089
Db 1444 LQEVCLRPSADPIFEPRQLPAFQEGIMSWSPKE-----KTYETDLAVLQWVEELEQRVI 1497
QY 1090 S-----LEKLXDFGECV----- 1101
Db 1498 MSDLIQIRGWTCPSPDSTREDLAYCEHLSDSQEDITWRGRGREGIAPQRKTNPLDLAVNR 1557
QY 1102 -IALQASVIKKFLOGFMAPKQRRKQLQESDAKTEEVDEEKKWVE-----EA 1147
Db 1558 LAALQONVERYLREPLWP-----THEVVLEKALLSTPNGAPGTTTETSY 1603
QY 1148 KVASALEKWKTAIREAOFTSRMHVLLGMLDACIKWDMSAENARCKVCPKGBDDKLIICD 1207
Db 1604 EITPRIVWRQTLERCRAAQVCLCLGQLERSIAWEKSVNKTCLVCRKGDNDDEFLLCD 1663
QY 1208 ECNKAFHLFCRLPALYEPDGEWQCPACQAPATARNRSGRNYTESASEDSDESDDEE 1267
Db 1664 GCDRGCHTYCHRPKMEAVPEGDWDFCTVC-----LAQQVGEGETQKPG 1705
QY 1268 EEEEEEEEDYEV---AGLRLRPKRTIRGKHSVILPPAARSGRRRPGKPKHSTRSQKAP 1324

Db 1340 LOWVELEQRVIMSLQIRGWTCPSDPSTREDLAYCEHLSDSQEDITWRGRGREGIAPQR 1399
 QY 1102 -----IALQASVIKKFLOGFMAPKQKRKLQSEDSAKTEVEDEKMWVE--- 1145
 Db 1400 KTTNPLDLAVMLAALAEQNVRRYLREPLWP-----THEVLEKALLSTPN 1445
 QY 1146 -----EAKVASALEKWKTAIREAQTFSRMHVLLGMLDACIKWMSAENARCKVCP 1195
 Db 1446 GAPEGTTEISYEITPRIRVWRMOTLERCSAAOVCLIGQLERSIAWKSVMNKVTCIVCR 1505
 QY 1196 KKGEDKLTICDNCKNAFLCLRPALVEVPDGEWQCACOPATARRNSRGRNYTESAS 1255
 Db 1506 KGDNDEFLLDCDGRGCHYCHRPKMEAVPEDWFCIVC-----LA 1547
 QY 1256 EDEDEDEDEDEDEDEDEDEDEYEV---AGLRLRPRTYIRGKHVIPPAAARGRRPGKK 1312
 Db 1548 QQVEGEFTQKPGFKRGKSGYSLNFSGDRRRRVLLRGRES--PAAGPRYSEGLS 1605
 QY 1313 PHSTRSQPKAPPVDDAEVDLVLQTKSSRRQSLQKCEIILHKIVKYRFSWPFEPV 1372
 Db 1606 PSKRRR-----LSMRNHSDLTFCETILMEMESHDAAPFLBFPV 1644
 QY 1373 TRDEAEDYVDVITHPMDFOTVONKSCGYSRVSQVDFLTDKQVFTNAEVYNCRGSHV 1429
 Db 1645 NPLRVSGYRIIKNPMDFSTMRLLRGYTSSEEFADALLVDFNCOTFNEEDSEV 1701
 RESULT 41
 AAM40239
 ID AAM40239 standard; protein; 1878 AA.
 XX AAM40239;
 AC AAM40239;
 DT 22-OCT-2001 (first entry)
 XX Human polypeptide SEQ ID NO 3384.
 DE Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia.
 XX Homo sapiens.
 OS
 XX WO200153312-A1.
 PN 26-JUL-2001.
 PD 26-DEC-2000; 2000WO-US034263.
 PF 23-DEC-1999; 99US-00471275.
 PR 21-JAN-2000; 2000US-00488725.
 PR 25-APR-2000; 2000US-00552317.
 PR 20-JUN-2000; 2000US-00598042.
 PR 19-JUL-2000; 2000US-00620312.
 PR 03-AUG-2000; 2000US-00653450.
 PR 14-SEP-2000; 2000US-00662191.
 PR 19-OCT-2000; 2000US-00693036.
 PR 29-NOV-2000; 2000US-00727344.
 XX (HYSE-) HYSEQ INC.
 PA Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;
 PI Zhou P, Goodrich R, Drmanac RT;
 XX WPI; 2001-442253/47.
 DR N-PSDB; AAI59395.
 XX Novel nucleic acids and polypeptides, useful for treating disorders such
 PT as central nervous system injuries.

XX Example 5; SEQ ID NO 3384; 10078pp; English.
 PS The invention relates to human nucleic acids (AAI57798-AAI61369) and the
 XX encoded polypeptides (AAM38642-AAM42213) with nootropic,
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: Immune system suppression,
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemia and
 CC C.N.S disorders. Note: The sequence data for this patent did not form
 CC part of the printed specification
 XX Sequence 1878 AA;
 SQ
 Query Match 6.6%; Score 528.5; DB 4; Length 1878;
 Best Local Similarity 19.4%; Pred. No. 7.8e-27;
 Matches 267; Conservative 178; Mismatches 427; Indels 505; Gaps 43;
 QY 415 QKSTGNSKSPKGLKTPKTKMQLTLDMAKGTQ---KMTAPRNSGGTPTTSKPKHKL 471
 Db 619 QAITGKGRPR---NTEKAKTKVEPKVGRGRPPKVKITEL-----LNKTDNRPLKKL 669
 QY 472 PPAALHLIAYVYKKNKREDKRSALSCVISTKAR-----LLSSDDRARLPBEL 518
 Db 670 EA-----QETLINEEDKAK-----IAKSKKWRQKQVQEGCOTTQGAARNKQKQET 715
 QY 519 RSLVQKRYELLEHKRWASMSSEORKEYLKKKREELKKLKEKAKER---REKEML----- 571
 Db 716 KSLAQK-----EAKKK-----SKAEKKGKTKQEKLEKVKREKEKVKMEKEBEVTKAK 765
 QY 572 -----ERLEKQKRY-----EDELGTG-KNLPAPRLVDTPEG--LNT 605
 Db 766 PACKADKTLATQRRLLEERQOQMILEEMKKTEDMCLTDHQLPDPFSRV---PGHLLPSG 822
 QY 606 LFGVAMVVFELSCVSGLLPDAQYPTAVS-LMEALSADKGGFLYLNRLVILLQTLQ 664
 Db 823 AFSCLTIVFELHSGVGLGFDPAKDVPSGLVQEGLLCQDGLSGEVDLVLRLKAAHL 882
 QY 665 TLQDETAEDYGEELGMKLSIPLTHSVSELVRLCLRRSDVQSESESGSDTDNKKDSAAPE 724
 Db 883 DPGFPSYCSQLKILGEKVSEIPLTRDNVSEILRCFLMAYG----- 922
 QY 725 DNEVQDEFLEKLETSEFPFELTSEBKLIQLITLALCHRI-----LMTYSVQDHMETROQMSAEL 780
 Db 923 ---VEPALCDRLRTQPPQAQPPQKAAVLAFLVHVLNGSTLIINEIDKTLSEMSSSYRNKK 979
 QY 781 W--KERLAVLKEENDKKRAEKQKREMEAKKNGKVBNGKGLKTDKTKRIIVKFEQVDOTE 838
 Db 980 WIVEGRLRLK---TVLAKTGTGSEVENEGP-----EELGRRRSSRIMETSMEWEE 1030
 QY 839 AEDMISAVKSRRLAIQAKEREIQEREMKV-KLERQAEERIRKHKAAAEKAFQEGIAK 897
 Db 1031 EESIAAVPGRR---GRRDGEVDATASSIPELERQIEKLSKR-----QLFPRKLLH 1079
 QY 898 AKLVNRTPIGTDNRNHNRYWLFSDVEVPGLPFTB-----KGWVHDSID 938
 Db 1080 SSQMLRAVSLGQDRYRRYRWL-PYLAGIFVEGTGNLVPEVVIKETEKTDLSUKVAHASLN 1138
 QY 939 YRF-----NHHCKDHTVSGDEDYCPRSK----- 961
 Db 1139 PALFSMKMELAGSNVTASSPARARGPRKTPGSMQPRHLKS-PVRGQDSEQPOALOPE 1197
 QY 962 -----KANLGKN-----ASMNTQHTGATE 980
 Db 1198 AQLHAPAQPPQLQLQLOQSHKGFLEQEGSPLSLGQSHDLSQSAFLSMLSTQSHSLLS 1257

Db 649 WIVEGRRLRLK-----TVLAKRTGRSEVEMEGP-----EECLRR-RSSRIMEVTSGMBEE 698
Qy 838 EAEDMISAVSRLLAIQAKKEREIQEREMKV-KLERQAEERIRKHAABAKAFQEGIA 896
Db 699 EEEESIAAVPGR-----GRRDGEYDATTASSPELERQIEKLSKR-----QLFFRKLL 747
Qy 897 KAKLVMRTPIGTDNRNRYMLFSDEVPLFTIE-----KGWVHDSI 937
Db 748 HSSQMLRAVSLGQDRYRYWVL-PYLAGIFVEGTEGNLVEEVTKKETDSLKVAHAHL 806
Qy 938 DYRF-----NHHCKDHTVSGDEYCPRSK--- 961
Db 807 NPALFSMKVELAGSNTTASSPARARGRPRKTKPGSMQPHLKS-FVRGDSGEQPOAQLOP 865
Qy 962-----KANLGN-----ASNTQHGAT 979
Db 866 EAQLHAPAQPOQLQLQSHKHGFLEQEGSLSLGQSQHDLSQSAFLSWSLQTSQSHSL 925
Qy 980 EVAVETTPKOG-----QNLWFLCDSQ----- 1001
Db 926 SSVLTPDSSPGKLPAPSPQPEEPDEARSSPDQALWFNISAQMPCNAAPTPTPPAVS 985
Qy 1002----- 1001
Db 986 EDQPTSPQOLASSKPMRNPAAANPCSPVQFSSTPLAGLAPKRRAGDPGEMPQSPGLGQ 1045
Qy 1002-----KELDELINCLHPQIRES 1019
Db 1046 PKRRGRPFSKFFKQMEORYLTOLTAQVPVPEMCSGMMWIRDPFEMDLAMLKALHPRGIREK 1105
Qy 1020 QKLEKRYODIHSIHL--ARKENLGLKSGDNQOE--LLNFRSLDLIEVATRLQKGLG 1076
Db 1106 ALHKHLNK-HRDFLOEVLRSADPIFEPRQLPAFQEGIMSPKE-----KTYETDLA 1158
Qy 1077 YVEETSEFEARVIS-----LEKLDGFCV----- 1101
Db 1159 VLQWVEELEQRVIMSDLOIRGWTCPSPDSTREDLAYCEHLSDSQEDITWRGREGGLAPQ 1218
Qy 1102-----TALQASVTKKFLQGFMAPKQKRKLQSEDSAKTEVDEEKQWE-- 1145
Db 1219 RKTNPDLAVMLAALQNVERRYLRLPWP-----THEVLEKALLSTP 1264
Qy 1146-----EAKVASALEKWKTAIREAQTFSRMHVLLGLMDACIKWDMGAENARCKVC 1194
Db 1265 NCAPGTTTETSYETPIRIVWRQTLRCRSAAQVCLCLGLQLERSIAWEKSNVKTCLVC 1324
Qy 1195 PKGDEDDKLIILDECNKAFHLFLCLRPALYEPDGEWQCPACQAPATARNRSGRNYTERSA 1254
Db 1325 RKGDNDEFLLDCGDRGCHIYCHRPKMEAVPEGDFCTVC-----L 1366
Qy 1255 SEDSDDESDEEEEEEEDYEV---AGLRPRKTRIGKHSVTPPAARSGRRPGK 1311
Db 1367 AQQVEGEFTQKPGFPKQKRSYSLNFSSEGDRRRVLLRGRES---PAAGPRYSEGL 1424
Qy 1312 KPHSTRSQPKAPPVDDAEVDDELVLQTKRSRRQSLQKCEIILHKIVKRFSPRRP 1371
Db 1425 SPSKRR-----LSMRNHSDITFCFIIIMESHDAAPFLEP 1463
Qy 1372 VTRDAEDYDVITHPMDFTQVQKSCSGSYRSVQEFITDMKQVFTNAEVYNCRSHV-- 1429
Db 1464 VNPRVSGYRRIIKNPMDFSTMRLLRGYTSSEFAADALLVFDNCQTFNEDDSEVGK 1523
Qy 1430-----LSCWVTEOCLVVLHKLPHGHPY 1453
Db 1524 AGHIMRRFFESRWEFFYQKQCSVQRQGWGVTLWHLPPPTFQTKC-----HFHLLMLFW 1578
Qy 1454 VERKRRKPPD 1463
Db 1579 VQTVRYNPD 1588

ID AAW81170 standard; protein; 1878 AA.
XX
AC AAW81170;
XX
DT 05-MAY-1999 (first entry)
XX
DE Human BAZ2-alpha protein.
XX
KW Transcriptional regulator; transcription; BAZ1-alpha; bromodomain; BAZ;
KW atypical zinc finger; testis; human; tumour; BAZ1-beta; BAZ2-alpha; drug;
KW BAZ2-beta; treatment; cancer; proliferative disorder; screening.
XX
OS Homo sapiens.
XX
FN WO9847920-A1.
XX
PD 29-OCT-1998.
XX
PF 17-APR-1998; 98WO-JP001783.
XX
PR 18-APR-1997; 97JP-00116570.
PR 24-OCT-1997; 97JP-00310027.
XX
PA (CHUS) CHUGAI RES INST MOLECULAR MEDICINE INC.
XX
PI Jones MH;
XX
DR WPI; 1998-583603/49.
DR N-PSDB; AAV68396.
XX
PT Transcriptional regulator gene family containing bromodomain - may be
PT expressed in testis tissue and is useful for treatment of cancer and
PT other proliferative disorders.
XX
PS Claim 1; Page 72-88; 187pp; Japanese.
XX
CC This sequence represents the human BAZ2-alpha protein, a member of a
CC family of transcriptional regulator genes containing a bromodomain (BAZ,
CC Bromodomain with Atypical Zinc finger) which are expressed specifically
CC in testis tissue and also in certain tumour lines. Transgenic cells may
CC be used for the preparation of the BAZ1-alpha, BAZ1-beta, BAZ2-alpha and
CC BAZ2-beta proteins. These proteins can be used in the treatment of cancer
CC and other proliferative disorders, and in screening of compounds for
CC their binding ability to the expression products (e.g. for use as drugs
CC by modulation of transcriptional regulation)
XX
SQ Sequence 1878 AA;
Query Match 6.4%; Score 514.5; DB 2; Length 1878;
Best Local Similarity 19.3%; Pred. No. 7.3e-26;
Matches 263; Conservative 181; Mismatches 441; Indels 479; Gaps 42;
Qy 415 QKSTGNSKSPKGLKTPKTKMKTLLDMAKGTQ---KMTAPRNSGGTPTRTSSKPHKL 471
Db 619 QAITKGRGRPR---NTEKAKTKEVPKVKRGRGPPKVKITEL-----LNKTDNRPLKKL 669
Qy 472 PPAALHLIAYKENKDREKDSALSCVISTARLLSSEDRARLPEELSLVQKRVLELH 531
Db 670 EA-----OETLNEEDKAKIAKSKKMKQKQVQGECLTTIOGAQNRKQETKSLKH 720
Qy 532 KKRWASMEEQREYLVKKRBEELKKLKEKAKER---REKE----- 569
Db 721 KE--AKKSKAEKEKGTKQKLEKKEKVKKEKKEKEKEVEVTKACKACKADKTATQR 778
Qy 570 MLEKLEKQ-----KREYDEQELTG-KNLPAFLRVDTPEG--LPNTLFGDVAMVVEFLS 618
Db 779 RLEERQKQOMILEEMKKPTEDMCLTDHQLPDFSRV---PGLTLPFGAFSDCLTIVEFLH 835
Qy 619 CYSGLLLPDAQYPITAVS-LMEALSADKGGFLYLNRLVILLQTLQLLQDEIAEDYGE 677
Db 836 SFGKVLGFDPAKQVPSGLVQELGQSDSLGEVQDILLRLKALHDPGPPSYCQSLKI 895
Qy 678 LGMKLSLEIPLTHSVSELVRLCLRRSDVQEESESGSDTDNKNDSAAFEDNEVQDEFLEKLE 737

896 LGKRVSEIPLTRDNVSEILRCFLMAYG-----VFPALCDRLR 932
738 TSEFFELTSEKLOILTALCHRI-----LMTYSVQDHMETRQQMSAELW--KERLAVLKEE 791
933 TQFPQAOQPOQAVALFAPVHELNGSTLIINEIDKLTESMSYRKKNKWIVEGRRLRK-- 990
792 NDKKRAEKOKRMEAKKENGKVENGLGKTRKIRIVKFPQVDEABDMISAVKSRRL 851
991 --TVLAKTRGRSEVEM-----GRPECLGRRSSRIMEETSGWEEBEEESIAAPGRR- 1042
852 LAIQAKKEREIOERENKV-KLROABEERIRKHAAAEXAFQEGIAKAKLVNRRTPIGTD 910
1043 ---GRDGEVDATASSIPELEROIEKLSK-----QLFRKGLHSSQMLRAVSLGQD 1092
911 RNHNRYWFSDEVPLGFIE-----KGWVHDSID----- 938
1093 RYRRRYWVL-PYLAGIFVGTGNLVPREVIKETDLSLKVAAHSLNPALFSMKMELAGS 1151
939 -----YRFNHKCDHTVSGDEDYCPRSK----- 961
1152 NTTASSPARARSLKTKPGFMQRHFKS-PVRGQDSEQFOAQLQPEAQLHVPAPQPOL 1210
962 -----KANLGKN-----ASMTQHGTAATEVAVETTTKQG-- 991
1211 QLQLOSHKGFLEQBSPLSGOSQHDLSQSAFLSMLSTQSHSLSSVLTPDSPGKL 1270
992 -----QNLWFLCDSO----- 1001
1271 DPAPSQPPEPEDEAESFDLQAFWNISAOQPCNAAPTPLAVSEDOQTPSPQOLASS 1330
1002 ----- 1001
1331 KPMNRPSAANPCSPVQFSTPLAGLAPKRRAGDPCGEMPOSPTGLQPKRGRRPSPKFKQ 1390
1002 -----XELDELNCLHPQIGIRESQLEKRYQDII 1033
1391 MEQRYLTQTAQVPVPEMCGMWIIPDPEMLDAMLKALHPRGIREKALHKLHNK-HRDPL 1449
1034 HSIHL--ARKPNLGLKSCDGNQOE-LLNFLRSLDIEVATLQKGLGYVEETSEFARVIS 1090
1450 QEVCLRPSADPIFEPRQLPAFQEGIMSWSPK-----KTYETDLAVLQWVEELSQRVIM 1503
1091 -----LEKLKDFGECV----- 1101
1504 SLDQIRGWTCPSPDSTREDLAYCEHLSDSQEDITWRGPGREGIAPQKRTTNPLDLAVMEL 1563
1102 IALQASVIKKFLOGFWAPKQKRKLQSEDSAKTEEVDEEKNWVE-----EAK 1148
1564 AALEQNVKRRYLREPLW-----THEVLEKALLSTPNGAPEGTTTTSIVE 1609
1149 VASALEKWKTAIREACTFSRMEVLLGLMDACIKWDMSAENARCKVCPKKGDDKLILCDE 1208
1610 ITPRIRIMQTLQRCSSAAHVCLCGLHLERSIANEKSVMNKVTCVCRKGDNDFFLLCQD 1669
1209 CNKAFLHFLCLRPALYEPDGEWQCPACQATARRNSRGRNYTEESASEDSEDESEEE 1268
1670 CDRGCHYIHRPKMEAVPEGDFCTVC-----LAQOVEGEFTQKQCF 1711
1269 EEEEBEEDDYEV--AGLRLRPRTYIRKGVHVIIPPAASGRPRPKKPHSTRSRSPKAPP 1325
1712 PKRGQKRSKSGYSLNFGDGRRRVLLKGRES--PAA--GPR-----YSEERLSPSK-- 1759
1326 VDAAEVDVLVLTQKSSRRQSLQKCEILHKIKVYRFSWPFRPVTRDEADYDVT 1385
1760 -----RRRLSMRHHSDLTFCETILMESHDAAWPFIFBPNRLVSGYRRITK 1808
1386 HPMDFQVONKSCSGYSVQSFLLTDMKQVFTNAEYVNCRGSHV 1429
1809 NPMDFSTWRERLLRGYTSSEFAADALLVFNCTQTFNEDDSEV 1852

AAB93746

ID AAB93746 standard; protein; 475 AA.

XX AC AAB93746;

XX DT 26-JUN-2001 (first entry)

XX DE Human protein sequence SEQ ID NO:13405.

XX KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.

XX OS Homo sapiens.

XX PN EP1074617-A2.

XX PD 07-FEB-2001.

XX PF 28-JUL-2000; 2000EP-00116126.

XX PR 29-JUL-1999; 99JP-00248036.

XX PR 27-AUG-1999; 99JP-00300253.

XX PR 11-JAN-2000; 2000JP-00118776.

XX PR 02-MAY-2000; 2000JP-00183767.

XX PR 09-JUN-2000; 2000JP-00241899.

XX PA (HELI-) HELIX RES INST.

XX PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

XX PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX WPI; 2001-318749/34.

XX PR Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.

XX PS Claim 8; SEQ ID NO 13405; 2537pp + Sequence Listing; English.

XX CC The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAB92445 to AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent present invention

XX CC all of which are used in the exemplification of the

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

Db 43 ASLSQVFLHSLTDRSVIWSKILNARCKICFKKGDAENWVLCDGCRGHHTYCVRPKLK 102
 QY 1224 EYVDGEMQCPACQATARNRSGRNYTBESASEDSEDEEBE-----EEEEEBEEDY 1279
 Db 103 TVPEGDFCEPRKQRRSLSSRQPSLESDEVEDSMGGEDDVEDGDEEGQSEEBY 162
 QY 1280 EV-----AGLRLRPKTTI-----RGKHSVIPPAARSGRRPKPHSTRSQP 1321
 Db 163 EYEQDEDDQEBEERVSLPKRGRPVQLPVKTRGKLSFSRSGQQQEGRPYRSRQSTP 222
 QY 1322 K-----APPVD-----DAEV----- 1331
 Db 223 KTVSSKTRGRSLRKINSAPPTKSLRTASRSTRSHGFLQADVVELLSPRKRGRKS 282
 QY 1332 -----DELVLQT----- 1339
 Db 283 ANNPENSPNPNFRVIAAKSEQSRSVNIASKLSQESKRRCKRQSPSPVTLGR 342
 QY 1340 RSSRQSS--LELQKCEEILHKIVKRFSPFPTRDEADYDVTTHPMDPQTQVQKNC 1397
 Db 343 RSSGQGGVHELSAFEQLVVELVRHDDSWPFLKLVSKIQVDPDYDIKKPIALNTIREKV 402
 QY 1398 SCGSYRSVQEFLLTMKQVFTNAEVNCRGS 1427
 Db 403 NKCEYKLASEFIDDIELMFNSCFEYNPNT 432

RESULT 45

ADRO9529
 ID ADR09529 standard; protein; 524 AA.

XX ADR09529;
 AC ADR09529;

DT 04-NOV-2004 (first entry)

DE Human protein useful for treating neurological disease Seq 3035.

XX human; oligo-capping method; diagnostic marker; gene therapy;
 KW osteoporosis; neurological disease; Alzheimer's disease;
 KW Parkinson's disease; dementia; short memory; cancer;
 KW sense or motor function; emotional reaction; fear response; panic;
 KW osteopathic; neuroprotective; nootropic; antiparkinsonian; cytoskeletal;
 KW tranquilizer.

XX Homo sapiens.

XX EP1447413-A2.

XX 18-AUG-2004.

XX 12-FEB-2004; 2004EP-00003145.

XX 14-FEB-2003; 2003JP-00102207.

PR 09-MAY-2003; 2003JP-00131452.

XX (REAS-) RES ASSOC BIOTECHNOLOGY.

XX Isogai T, Yamamoto J, Nishikawa T, Isono Y, Sugiyama T, Otsuki T;
 PI Wakamatsu A, Ishii S, Nagai K, Irie R;

DR WPI: 2004-583265/57.

DR N-PSDB; ADR07573.

XX New 1995 cDNA, useful for treating osteoporosis, neurological diseases,
 PT Alzheimer's diseases, Parkinson's diseases, dementia and various cancers.

XX Claim 1; SEQ ID NO 3035; 2686pp; English.

XX This invention relates to novel, isolated full length human cDNA
 CC molecules and the encoded proteins thereof. Specifically, it refers to
 CC cDNA clones obtained by an oligo-capping method, where none of these
 CC clones are identical to any known human mRNAs. The present invention

CC describes an immunoassay to identify agonists and antagonists, as well as
 CC antibodies, antisense molecules and siRNAs that can all be used to bind
 CC to and modulate expression of the cDNA molecules. As such, these
 CC molecules are useful for diagnostic markers or therapeutic targets for
 CC the various diseases or morbid states. In particular, they are useful in
 CC gene therapy for treating osteoporosis, neurological disease, Alzheimer's
 CC disease, Parkinson's disease, dementia, short memory and various cancers,
 CC as well as for maintaining equilibrium of sense or motor function, and
 CC for treating emotional reaction, fear response and panic. Accordingly,
 CC they exhibit osteopathic, neuroprotective, nootropic, antiparkinsonian,
 CC cytoskeletal and tranquilizer activities. This polypeptide is a protein
 CC encoded by a full length human cDNA sequence of the invention. NOTE: This
 CC sequence is not given in the sequence listing of the specification but
 CC can be obtained on CD-ROM from the European Patent Office, Vienna Sub-
 CC office.

XX SQ Sequence 524 AA;

Query Match 4.6%; Score 368.5; DB 8; Length 524;

Best Local Similarity 22.5%; Pred. No. 1.5e-16;
 Matches 115; Conservative 78; Mismatches 186; Indels 131; Gaps 13;

QY 988 PKGQNLWFLCDSQKELDELNCLHPQIGRESOLKRLERYQDIHSIHL--ARKPNLG 1045

Db 52 PKMCSGWWIRDPDM-LDAMLKALHPRGIREKALHKLHKLK-HRDFLQEVCLRPSADPIFE 109

QY 1046 LKSCDGNQOE-LLNFLRSDLIEVATRLQKGGLYGYEETSEFEARVIS----- 1090

Db 110 PROLPATQEGIMSWSPKE-----KTYETDLAVLQWVEELQORVIMSDLOIRGWTCPSPD 163

QY 1091 -----LEKLDKDFGECV-----IALQASVTKKFLQG 1115

Db 164 STREDLAYCEHLSDSQEDITWRGRGREGAPQRKTTPDLAVMLAALQNVRYLRE 223

QY 1116 FMAPKQKRLQSDSDSAKTEEVDEBEKMWVE-----EAKVASALEKWKTAIRE 1162

Db 224 PLWP-----THEVLEKALLSTPNGAPEGTTTTEISYEITPRIRVNRQTLE 269

QY 1163 AQTSRMHVLLGMLDACIKWDMSAENARCKVCPKKGEDDKLLLCDECNKAFLFLCLRAL 1222

Db 270 CRSAQVCLGQLGQLERSIAWEKSVKVTCLVCRKGDNDDEFLLCDGCRGCHYCHRRPKM 329

QY 1223 YEVPDGEWQCPACQAPATARRNSGRNVTESASDESDEDEDEDEDEDEDEDEDEDEDE 1281

Db 330 EAVPEGDMFCTVC-----LAQVEGEFTQKPGFKRGKRSYSLN 371

QY 1282 --AGLRLRPKXIRGKHSVIPPAARSGRRPGKPHSTRSQPKAPPVDDAEVDLVLQTK 1339

Db 372 FSEGDGRRRRVLLRGRES--PAAGPRYSEGLSPSKRR----- 408

QY 1340 RSSRQSLLEKQCEEILHKIVKRFSPFPTRDEADYDVTTHPMDPQTQVQKSC 1399

Db 409 LSMRNHSDLTFCCEILMEMESHDAAWPFLEFVPNRLVSGYRRIRIKNPMDFSTRERLLR 468

QY 1400 GSYRSVQEFLLTMKQVFTNAEVNCRGSHV 1429

Db 469 GGYTSSEFAADALLVFNDCQTFNEDDSEV 498

RESULT 46

AAG03100

ID AAG03100 standard; protein; 60 AA.

XX AAG03100;

DT 06-OCT-2000 (first entry)

XX Human secreted protein, SEQ ID NO: 7181.

XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
 KW gene therapy; chromosome mapping.

OS Homo sapiens.

XX EP1033401-A2.
 XX 06-SEP-2000.
 XX 21-FEB-2000; 2000EP-00200610.
 XX 26-FEB-1999; 99US-0122487P.
 XX (GEST) GENSET.
 XX Dumas Milne Edwards J, Duclert A, Giordano J;
 XX WPI; 2000-500381/45.
 DR N-PSDB; AAC03106.
 XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
 PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures.
 XX Claim 13; SEQ ID NO 7181; 71pp + Sequence Listing; English.
 XX The present sequence is a polypeptide encoded by one of a large number of
 CC 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs were
 CC prepared from total human RNAs or polyA+ RNAs derived from 30 different
 CC tissues. EST sequences usually correspond mainly to the 3' untranslated
 CC region (UTR) of the mRNA because they are often obtained from oligo-dT
 CC primed cDNA libraries. Such ESTs are not well suited for isolating cDNA
 CC sequences derived from the 5' ends of mRNAs and even in those cases where
 CC longer cDNA sequences have been obtained, the full 5' UTR is rarely
 CC included. 5' ESTs are derived from mRNAs with intact 5' ends and can
 CC therefore be used to obtain full length cDNAs and genomic DNAs. 5' ESTs
 CC are also used in diagnostic, forensic, gene therapy and chromosome
 CC mapping procedures. They are used to obtain upstream regulatory sequences
 CC and to design expression and secretion vectors
 XX Sequence 60 AA;
 SQ
 Query Match 4.4%; Score 352; DB 3; Length 60;
 Best Local Similarity 100.0%; Pred. No. 9.6e-17;
 Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 902 MRTPTGTRNNRNYLWLSDEVPGLFIEKGWVHDSIDYRFNHHCKDHTVSGDEDYCPRSK 961
 DB 1 MRTPTGTRNNRNYLWLSDEVPGLFIEKGWVHDSIDYRFNHHCKDHTVSGDEDYCPRSK 60
 RESULT 47
 ABR52663
 ID ABR52663 standard; protein; 1264 AA.
 XX ABR52663;
 XX 20-JUN-2003 (first entry)
 XX Protein sequence #SEQ ID 191.
 XX Multiprotein complex; eukaryote; drug target; diagnosis.
 XX Saccharomyces cerevisiae.
 XX EP1258494-A1.
 XX 20-NOV-2002.
 XX 20-DEC-2001; 2001EP-00130253.
 XX 15-MAY-2001; 2001EP-00111774.
 XX (CELL-) CELLZOME AG.
 XX Bauer A, Gavin A, Grandi P, Krause R, Kruse UD, Kuester BD;
 PI Marzioch M, Schultz JD, Superti-Furga GD;

XX WPI; 2003-250078/25.
 DR N-PSDB; ACC60705.
 XX New isolated protein complexes useful for diagnosing a disease or
 PT disorder, or as a target for an active agent of a pharmaceutical,
 PT preferably a drug target in the treatment or prevention of disease or
 PT disorder.
 XX Disclosure; SEQ ID NO 191; 17pp + Sequence Listing; English.
 XX The invention relates to multiprotein complexes from eukaryotes. Proteins
 CC of the invention and DNA sequences encoding them are given in records
 CC ABR52568-ABR53903 and ACC60610-ACC61944 respectively. The complexes are
 CC obtainable by using a protein as a bait and isolating the set of proteins
 CC which is attached thereto from cells. Such protein complexes may comprise
 CC up to 30 distinct proteins. Protein complexes of the invention are useful
 CC for diagnosing a disease or disorder, or as a target for an active agent
 CC of a pharmaceutical, preferably a drug target in the treatment or
 CC prevention of a disease or disorder. Note: The sequence data for this
 CC patent is not represented in the printed specification, but is based on
 CC sequence information supplied by the European Patent Office. The complete
 CC document is available on CD-ROM
 XX Sequence 1264 AA;
 SQ
 Query Match 4.3%; Score 345.5; DB 6; Length 1264;
 Best Local Similarity 19.5%; Pred. No. 2e-14;
 Matches 271; Conservative 206; Mismatches 492; Indels 423; Gaps 58;
 QY 3 PLLGRKPLVNPPLGPEPFTTIPHTQAFRTREYERLERYSERIWTCKSTGSQLPH 62
 DB 8 PILLPDPKPL--PLDLNVQVWHIEBTGEWFSYEEFLERFDYTRHHFTCEITGTSCLTF 65
 QY 63 KEAWEEQEVAELKEEPANWEKLVLEWVHN--TASLEKLVDTAWLEIMTKYAVGECSD 121
 DB 66 FOALDSEETQFYVEDRFLKREPVARFLHFNIRLDALVEKYARFKNDFPFGE--- 122
 QY 122 FEVGKMKLVKIVKIHLEKVEDEATEKSKGACDSSDKDENS---QIADHQKKE 177
 DB 123 -----VVYLRKQKDSSTSSNQOSTPQDDWVEINSGVNGPGLPQYQRR 168
 QY 178 TVVDEDEGRRSINDRARRSPKLPSTLKKGERKWAPPFLPHKYDVKLQNEKDII SN-- 235
 DB 169 YVIREKVQFNATINPESREIVPAHT-----KY---MLIEEAASSNKS 208
 QY 236 --VPADSLIRTPPNKEIVRYFIRHNALRAGTGENAPVVEDELVKYKSLPSKSDFL 293
 DB 209 FIVDQGIYRDRSTFTKHLIKCFKKITLQIRASSKMGAPCVCVPEYLAMYGLTMEWPK--- 265
 QY 294 DPYKMTLNPSTKRK---NTGSPD-----RKPSKSKTDNSSLSPNPKLWCHVHL 342
 DB 266 DMLKYKEDEPVVARNSANVSSPESEKVKQSKSGKSNSTNSDASNKKETK--- 318
 QY 343 KKSLSGSPVKVKNKSNKSPPEHLEEMKQMSPNKLHTNFHIPKKGPAPKPGHSDKPL 402
 DB 319 ---KRKPTVNDSENNSEED-----KKKGQNVTS 347
 QY 403 KAKGRSGILNGQKSTGNSKSPKGLKTPKTKMQMTLL-----DMAKGTQKMTAPRNS 457
 DB 348 HSKRKKKE--ANBEPNTENVES---VPTPANABPQAVTITSIMDDLA----- 389
 QY 458 GGTPTRTSSKPHKHLPPAALHLIAVYKENKREDKRSALSCVISTKARLLSSDRARLPEE 517
 DB 390 -----LPYQH--PPNIFPNLTYYNEK-----LECI-----SLGSTKLSRPPDS 425
 QY 518 LRSLVQKRYELLE---HKRWASMSSEQRKEYLK-----KXREEL 554
 DB 426 FGKLLQ--AYQFLNTFGSKIQLSHFSLDQFITSKCTDPVELKGEVVLVNVIRNRTQTSKEQEI 484
 QY 555 KKK---LKEKAKERREKEMLERLEKQRYEQBELTGK---NLPAPRLV--DPEG---LPN 604
 DB 485 ENNGLPKMKNAETTTEDSENPSQWRNSFRIDMIMKRNDSKVYKIVHDDPASDDILDN 544


```
Db 348 HSKRKKE-ANBEPNTVENES-----VPTPANAEPQAVTITSIMDDLA----- 389
Qy 458 GCTPTSSPHGLHPAALHLIAYKENKDRKDSALSCVSKTARLLSSSDRARLPPE 517
Db 390 -----LPYQH-PNIIFNPTTYNEK-----LECI-----SLGSTKLRPPDS 425
Qy 518 LRSLVQKRYELLE--HKKRWASMSBEQRKEYLK-----KREEL 554
Db 426 FGKLLQ-AVQFLNTFGSKICLSHFSLDQFITSLKCTDPVELKGEVVLNIRQTSEKEI 484
Qy 555 KKK---LKKAKERREKEMLEKQRYDEDELTKG---NLPAFLV-DTPEG---LPN 604
Db 485 ENNGLPKMKNAETTEEDSENPSDQWNSFIRMIMKRNDSKVEYKIVHDDPASDDILDN 544
Qy 605 TLFGDVAMVVEFLSCYSGLLLPDAQVPIITAVSLMEALSADKGFYLNVRVLILLQTLQ 664
Db 545 INHNGSALLIEVFTALLRIFI-NEEGDWCICIVENWIIDKG-----VLMER--- 590
Qy 665 TLLQDEIADYGEGLMKLSEIPLTLHVSSELVRL--CLRR--SDVQEESEG-SDTDONKD 719
Db 591 ---KDERGE--GEAKQKNAHYFLODKKEKIDNLKDTLAKENATEVQKESDAKNEINSESD 645
Qy 720 SAAFDNEVDREFLEK-----LETSEEFELTSE-- 747
Db 646 SKSDSDSEERDPKLEKCLNRYNVNWIERTLKRQFNNSYWLIIILGLVLEDCRHLPMYTEFI 705
Qy 748 ---EKL-----QILTALCHRI-----LMTY-----SVQDHM 770
Db 706 DSFIEKIIPKOISATQLPKQLWRNFCRLFSKDNVNALWILVDLSHFSPDIIKAAVDDSM 765
Qy 771 ETROQMSAELKERVLALEENDKKRAEKQKEMEA-----KNKENGKVENGLKGTDR 824
Db 766 E---LCQIRSERPKVAELKTEAAVLNLOGDLQAIQKLNKTDTENTPSADGDKD- 820
Qy 825 KRIIVKPEQVDTAEADMTSAVSKRRLLAIQAKREIQREREMKVKLERQABEERIRKH 884
Db 821 -----DSESNSEPIDLIIIEKKQKLIBEODKQVQALQ----- 852
Qy 885 AAQKAFQSGIAKAKIWMERTPIGTDRNHRV-WLFSDEVP-----GLFIEKGWHDQSI 937
Db 853 ---SDKNFLNCLFENDLQKLKPLGLDRYGNRYFWLDHNGVPPPPQYAGM-NETPKSNNSL 909
Qy 938 DY-----RPNHCKDH-----TVSGDEDY 956
Db 910 SYHSGRLLIQPKASSAKFFLVNDSQLSNQWKIRNSEGISSEATREVFGISKTSGSYNY 969
Qy 957 CPRSKKAN-LGNKASNTO-HGTATEVAVETTPKQ---GQNLWFLCDSQKELDELLNCL 1011
Db 970 VENGIEVELLSDNDRVNPILIELTPQKIMDETSPGRLLLSPPQWYCIDKLEDLSRIMDWL 1029
Qy 1012 HPQGIRESOLKERLEKRYQDIITHSLARKPNLGLKSCDGNQELL-----NPLR 1060
Db 1030 DNWGRKHDLLQOIRIMERIKSSLSL-RDHALSUTAFTKNBEKLLKELENNEFTENELN 1088
Qy 1061 SOLIEVATRLQGGGLGYVEETSEFARVLSLEKLDQFGECVIALQASVTKFLQFMAPK 1120
Db 1089 VDSMDVDDK--NSGV-----KSEVDVQDAEEK---REAVIDEKLEVIADL----- 1130
Qy 1121 QKRKLQSEDSAKTEV-----DEEKQVVE-----AKVASALEKWKTAIR 1161
Db 1131 -----MKLDDSKTRNVNLRIQEBDQORDELLEQKKSIINSRQPGARILARSEKRTKIS 1185
Qy 1162 EAQTFSRMHWLL 1173
Db 1186 RGNKVNKQTEIL 1197
```

RESULT 49
ABB58514
ID ABB58514 standard; protein; 2897 AA.
XX
AC ABB58514;

```
XX 26-MAR-2002 (first entry)
XX Drosophila melanogaster polypeptide SEQ ID NO 2334.
XX Drosophila; developmental biology; cell signalling; insecticide; pharmaceutical.
XX Drosophila melanogaster.
XX WO200171042-A2.
XX 27-SEP-2001.
XX 23-MAR-2001; 2001WO-US009231.
XX 23-MAR-2000; 2000US-0191637P.
XX 11-JUL-2000; 2000US-00614150.
XX (PEKE ) PE CORP NY.
XX Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-656860/75.
XX N-PSDB; ABL02617.
XX New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signaling and cell-cell interactions.
XX Disclosure; SEQ ID NO 2334; 21pp + Sequence Listing; English.
XX The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-ABB72072). The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX Sequence 2897 AA;
Qy Query Match 4.3%; Score 342; DB 4; Length 2897;
Best Local Similarity 19.0%; Fred. No. 1.1e-13;
Matches 265; Conservative 182; Mismatches 442; Indels 504; Gaps 55;
Qy 61 THKEAWEE---EQSVAELLKEEFPAPWYEKLVLEVMVHHTASLEKLVDTAWLEIMTKYAV 116
Db 603 TESEAIEGVRNRTVSVLQNNNSDSSITQDQHKVPADLNN-----TKKPV 652
Qy 117 GEEDCFVGVGKMKLVKIVKHPLEKVEDEATEKKSGDGDSPSSDKSNSQIAQDHQKK 176
Db 653 TEKCVLESG---TVEDKLVINH-----QOISTNALDTACDEKLS-CETESPVPNQHDSK 702
Qy 177 ETVVKED-EGRESINDRARRSPKLPKLSLKGKRWAPPKLPKHQYDVKLQNEDKIISN 235
Db 703 MKSEQTATSAKESVSESS-----PTD---GEDETAKSKCLIDFYAGTKNDKGIINNK 752
Qy 236 V-PADSLIRT--ERPKNKEIVRYFIRHNALRAGTGENAPWVVEDELVKKYLSPSKFSDFL 292
Db 753 IDKAEGVKTLLODDPPKDKSKSEINTTCLINISPSAEHRISE----- 796
Qy 293 LDPYKMYTLNPTSKRK-----NTGSPDRKPSKSKTDNSSLSPNPKLVCHVHLKKS 345
Db 797 ---KVQTEPSTSQNLLFEDNGKSGEVDGKSRSTGAVEEISKSTL-----LNRKR 845
Qy 346 LSGSPKLVKNSKNSKSPHEHLEEMMKMSPNKLHNFHPKPGPAKPKGKHSKSDPLKAK 405
Db 846 LNDQSALRNS--TSESEVHEE-----PQDDPTLDDLVDYGGKRIRKVR 887
```

QY 406 GRSKILNGQSTGNSKPKGLTKPTKTKMKTLLDMAKGTQKMTAPRNSGCTPTRSS 465
Db 888 PK-----TTNAESRRK-VEAQKQIETT-----SSGEDDARIRKCI 925
QY 466 KPHKHLPPAALHLIAYYKENDREDKRSALSCVISKTARLLSSDRARLPBLSLVQR 525
Db 926 APHTKTKTELEII-----ERLKKSI-----EMDLPEKTKVEM 962
QY 526 YELLEHHKRW-----SWSERQKBYLKKREELKKLKEKAKERREKEMLERLEKQRY 580
Db 963 PTLMQKTREATPTPIISLSPQKKTPTKPLKKNLLTQLRQ---EBSEBETIPR---KRT 1015
QY 581 EDQELTGKNLPAFLVDTPEGL-----PNTLFGDVAMVVEFLSCVSGLLLPDAQYP 631
Db 1016 NSETLV-PAIPASNVLCOPDERHRKRRSSEDANEAFSKESPTE-----VP 1060
QY 632 ITAVSLMEALSADKGGFLYLRNVLVILLQTLLODEIAEDYGGELGWLKLSIPLTLHS 691
Db 1061 PSAVS--EKLKRN-----EQDIQEV-----EDPLAM-S 1087
QY 692 VSELVRLCLRRSDVOEES-----GSDTDDNKDSAAAFEDNEVQDEFLEKLETS 739
Db 1088 VKDSLRSKADQSPVPEGSARRSGRGPAVHSELPOPKRTGGGARDKMQPEVNAELKQ- 1146
QY 740 EFFELTSEKQLLTALCHRILMTYSVODHMETRQMSAELWKLAVLKEENDKRAEK 799
Db 1147 ---ESEDDEKI-----STKIKSEAKDDPAPE-----SPENRKKPEEK 1180
QY 800 QKRKEMEAKNKGKENGVLGKTKDKRIVKPEPOVDT---EADMTISAVKSRRLLAIQ 855
Db 1181 PIKEE---PNEEPKPKVGRGPKKR-----EVDTTNIIETNDSETPVQSRRIAQQ 1230
QY 856 AKKEREIQEREMKVKLEQAEERIRKHAAAEKAFQEGIAKAKLVMRRTPIGTDNRNR 915
Db 1231 KIKEAEARRKQBEVALRTWKQE--LKKKKKAEKA----- 1263
QY 916 YWLFSDVPGULFIEKGWHDSDIDYFNHCKDHTV---SGDEDYCPKSKKANLGNKSMN 972
Db 1264 ---DPTVLEPSGEE----- 1274
QY 973 TQHGTAIVAVETTPKQ---GONLWFLCDSOKELDELLNCLHPQGIRESOLKERLEKR 1028
Db 1275 -SESASAEAEARNKKKKCPGKDGW-SSDSEEQ-----PESEEEEEPPH 1319
QY 1029 YODIISHLARKPNLGLKSCDGNQELNFRSLDIEVATRLQKGLGYVBETSEFEARV 1088
Db 1320 YE-----TDPGSPF---FRSD-----HEFSPE- 1338
QY 1089 ISLEKLDGFCVIALQASVIKKFLOGFWAPKQKRLQSEDSAKTEEVDEBEKMWVEAK 1148
Db 1339 ---SELEDESQV-----PMKRARTVRKENADDEEDAE----- 1371
QY 1149 VASALEKNTAIREAQTFSRMHVLGLMDACIKWDMSAENARCKVCPPKKGEDDKLILCD- 1207
Db 1372 -----ACQCKGSDHPWILLCDT 1390
QY 1208 -ECNKAFLFCLRPALYVPGEWQCPACQ-----PATARNRSGNRYTE 1251
Db 1391 PTCNKGYHCSCLSPLVFIPEGDWHCPQCQBOLIAALERQLEQYDVTVAQKQBRILAE 1450
QY 1252 ESASEDSEDES-----DEEEEEEEEEDEYEVAGLRPRKTIIRKHSVIPPAAARS 1305
Db 1451 EOAEEROLEAATLAADENPKSEKEDDRDDMAVGKAEKVKRRRGDGRINRRAAKR 1510
QY 1306 G--RRPGKKPHSTR------SOPKAPPVDDAEVDELVLQTKRSSRSQSLE 1348
Db 1511 GTRRRRGNESDSHRKSLGSGSRSGSDSDNSTSPSDSDDEPIYKLRKRRQINVSyr 1570
QY 1349 LQKCEIILHKIVK 1361
Db 1571 LNEYDDLINSALK 1583

RESULT 50
AAM14835
ID AAM14835 standard; protein; 65 AA.
XX
AC AAM14835;
DT 12-OCT-2001 (first entry)
XX
DE Peptide #1269 encoded by probe for measuring cervical gene expression.
XX
KW Probe; human; microarray; gene expression; cervical epithelial cell;
KW cervical cancer.
OS Homo sapiens.
FN WO200157278-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US000670.
XX
XX 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-488901/53.
XX
XX Human genome-derived single exon nucleic acid probes useful for analyzing
XX gene expression in human cervical epithelial cells.
XX
XX Claim 27; SEQ ID NO 19661; 487pp; English.
XX
XX The present invention relates to human single exon nucleic acid probes
XX (SENP; see AAI10068-AA128459). The present sequence is a peptide encoded
XX by one such probe. The SENPs are derived from human HeLa cells. The SENPs
XX can be used to produce a single exon microarray, which can be used for
XX measuring human gene expression in a sample derived from human cervical
XX epithelial cells. By measuring gene expression, the probes are therefore
XX useful in grading and/or staging of diseases of the cervix, notably
XX cervical cancer. Note: The sequence data for this patent did not form
XX part of the printed specification, but was obtained in electronic format
XX directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 65 AA;
XX
Query Match 4.3%; Score 341; DB 4; Length 65;
Best Local Similarity 100.0%; Pred. No. 6.2e-16;
Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 233 ISNPADSLIRTERPPNKEIVRYFIRHNALRAGTGENAPWVVEDELVKYSLPSKFSDFL 292
Db 1 ISNPADSLIRTERPPNKEIVRYFIRHNALRAGTGENAPWVVEDELVKYSLPSKFSDFL 60
QY 293 LDPYK 297
Db 61 LDPYK 65
Search completed: March 9, 2005, 14:07:15
Job time : 154.963 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 9, 2005, 13:54:12 ; Search time 34.6973 Seconds
(without alignments)
4242.638 Million cell updates/sec

Title: US-10-702-148-21
Perfect score: 10138
Sequence: 1 MGQTKSTSSGGNRKCNQEQ.....AGHNRKYEKKWDTFKVS 1972

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 65 summaries

Database : Issued Patents_AA.*
1: /cgn2_6/ptodata/1/1aa/5A_COMB.pep.*
2: /cgn2_6/ptodata/1/1aa/5B_COMB.pep.*
3: /cgn2_6/ptodata/1/1aa/6A_COMB.pep.*
4: /cgn2_6/ptodata/1/1aa/6B_COMB.pep.*
5: /cgn2_6/ptodata/1/1aa/PTCUS_COMB.pep.*
6: /cgn2_6/ptodata/1/1aa/backfiles.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	10138	100.0	1972	4	US-09-418-710-21
2	10138	100.0	1972	4	US-09-839-479-21
3	10022.5	98.9	1969	4	US-09-418-710-72
4	10022.5	98.9	1969	4	US-09-839-479-71
5	2493.5	24.6	1876	4	US-09-418-710-71
6	2493.5	24.6	1876	4	US-09-839-479-70
7	2482.5	24.5	1878	4	US-09-418-710-13
8	2482.5	24.5	1878	4	US-09-839-479-13
9	888	8.8	176	4	US-09-418-710-66
10	888	8.8	176	4	US-09-839-479-65
11	726.5	7.2	1674	4	US-09-418-710-1
12	726.5	7.2	1674	4	US-09-839-479-1
13	709	7.0	1673	4	US-09-418-710-70
14	709	7.0	1673	4	US-09-839-479-69
15	645.5	6.4	800	4	US-09-270-767-45282
16	538.5	5.3	1525	4	US-09-418-710-69
17	538.5	5.3	1525	4	US-09-839-479-68
18	536	5.3	1540	4	US-09-949-016-7037
19	533	5.3	1531	4	US-09-418-710-29
20	533	5.3	1531	4	US-09-839-479-29
21	532	5.2	1527	4	US-09-418-710-27
22	532	5.2	1527	4	US-09-839-479-27
23	445.5	4.4	175	4	US-09-418-710-65
24	445.5	4.4	175	4	US-09-839-479-64
25	406	4.0	196	4	US-09-690-454-94
26	390	3.8	2781	4	US-09-698-295-10
27	378.5	3.7	3259	4	US-09-949-016-6507

28	372.5	3.7	2733	4	US-09-949-016-11433	Sequence 11433, A
29	368	3.6	2907	4	US-09-698-295-1	Sequence 1, Appl
30	362.5	3.6	8991	4	US-08-714-741-32	Sequence 32, Appl
31	357	3.5	1564	4	US-10-144-198-2	Sequence 2, Appl
32	357	3.5	1564	4	US-10-144-198-4	Sequence 4, Appl
33	327	3.2	1240	4	US-09-538-092-658	Sequence 658, App
34	326.5	3.2	1162	2	US-08-728-323A-2	Sequence 2, Appl
35	326.5	3.2	1162	3	US-09-298-568-2	Sequence 2, Appl
36	326.5	3.2	1162	4	US-09-410-399-2	Sequence 2, Appl
37	326.5	3.2	1162	4	US-09-894-273-2	Sequence 2, Appl
38	320	3.2	2375	4	US-09-538-092-1131	Sequence 1131, Ap
39	316.5	3.1	3210	4	US-09-538-092-1154	Sequence 1154, Ap
40	316.5	3.1	3248	1	US-08-353-700-1	Sequence 1, Appl
41	316.5	3.1	3248	5	PCT-US95-16216-1	Sequence 1, Appl
42	310	3.1	919	4	US-09-949-016-6954	Sequence 6954, Ap
43	310	3.1	2954	4	US-09-150-867-1	Sequence 1, Appl
44	309.5	3.1	1898	4	US-08-056-200-94	Sequence 94, Appl
45	309.5	3.1	1898	2	US-08-800-644-94	Sequence 94, Appl
46	309.5	3.1	1898	4	US-09-538-092-1280	Sequence 1280, Ap
47	305.5	3.0	2842	1	US-07-741-940-7	Sequence 7, Appl
48	305.5	3.0	2842	1	US-08-289-548A-7	Sequence 7, Appl
49	305.5	3.0	2842	1	US-08-452-654-7	Sequence 7, Appl
50	305.5	3.0	2842	4	US-08-449-731-7	Sequence 7, Appl
51	305.5	3.0	2843	1	US-08-452-655B-2	Sequence 2, Appl
52	305.5	3.0	2843	1	US-08-452-655B-7	Sequence 7, Appl
53	305.5	3.0	2843	3	US-08-450-582-2	Sequence 2, Appl
54	305.5	3.0	2843	3	US-08-450-582-7	Sequence 7, Appl
55	305.5	3.0	2843	4	US-10-092-138A-30	Sequence 30, Appl
56	305.5	3.0	2843	4	US-09-538-092-1007	Sequence 1007, Ap
57	305.5	3.0	2973	2	US-08-821-355A-7	Sequence 7, Appl
58	305.5	3.0	2973	2	US-09-003-687A-7	Sequence 7, Appl
59	305.5	3.0	2973	3	US-09-136-605-7	Sequence 7, Appl
60	302.5	3.0	3878	4	US-09-914-259-11	Sequence 11, Appl
61	302	3.0	1402	4	US-09-248-796A-14503	Sequence 14503, A
62	301.5	3.0	2662	4	US-09-595-684B-31	Sequence 31, Appl
63	301.5	3.0	2843	1	US-07-741-940-2	Sequence 2, Appl
64	301.5	3.0	2843	1	US-08-289-548A-2	Sequence 2, Appl
65	301.5	3.0	2843	1	US-08-452-654-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1
US-09-418-710-21
; Sequence 21, Application US/09418710
; Patent No. 6596482
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042001
; CURRENT FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: US/09/418,710
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 21
; LENGTH: 1972
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-418-710-21

Query Match 100.0%; Score 10138; DB 4; Length 1972;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1972; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MGQTKSTSSGGNRKCNQEQSKQPLDARVDIKDKKPKKAMESSSNSDSGTSSTDS 60
|||||

Db 1 MGQTKSTSSGGRKCNQEQSKNQPLDARVDKI KDKKPKKXAMBESSNSDSGTSSDTS 60
Qy 61 SEG1SSSDSDLDLEDEDEEDDSQSI BESEDDDDSDSEAOHKSNNQVLLHIGISDPKADGQKA 120
Db 61 SEG1SSSDSDLEDEDEEDDSQSI BESEDDDDSDSEAOHKSNNQVLLHIGISDPKADGQKA 120
Qy 121 TEKAQEKRIHOPLPLAFESQTHSFQSQOKQPOVLQSQQLPFI FQSSQAKEESVNXHTTSVIQ 180
Db 121 TEKAQEKRIHOPLPLAFESQTHSFQSQOKQPOVLQSQQLPFI FQSSQAKEESVNXHTTSVIQ 180
Qy 181 STGLSVSNKPLSLVNQAKKETYMKLI VPSPPVLKAGNKNTSESSLLTSELRSKREQYKQ 240
Db 181 STGLSVSNKPLSLVNQAKKETYMKLI VPSPPVLKAGNKNTSESSLLTSELRSKREQYKQ 240
Qy 241 APPSOLKQESSKSLKVVIAALSHPKATSSSPAHPKQTLNHNHPFIFTNALLGNHQPNG 300
Db 241 APPSOLKQESSKSLKVVIAALSHPKATSSSPAHPKQTLNHNHPFIFTNALLGNHQPNG 300
Qy 301 VIQSVIQEAPLALTTKTKWQSKINENIAAASSTPFSPPVNLSTGRRTPGNOTPWPSPAS 360
Db 301 VIQSVIQEAPLALTTKTKWQSKINENIAAASSTPFSPPVNLSTGRRTPGNOTPWPSPAS 360
Qy 361 PILHSQKKEKAVSNVNPVKTHSHHPAKSLVEQFRGTDSDIPSSKDSSEDSNEDEEDE 420
Db 361 PILHSQKKEKAVSNVNPVKTHSHHPAKSLVEQFRGTDSDIPSSKDSSEDSNEDEEDE 420
Qy 421 BEDEDEDDSDSDSQSSDSNSSDTEGSEEDDDDDKQDESDDTEGEKTSMKLNKTT 480
Db 421 BEDEDEDDSDSDSQSSDSNSSDTEGSEEDDDDDKQDESDDTEGEKTSMKLNKTT 480
Qy 481 SSVKSPMSLTHGHPNHLHAKPGSAPALCSQSQSPAFILGTSSTSLTSSPHSGTSKR 540
Db 481 SSVKSPMSLTHGHPNHLHAKPGSAPALCSQSQSPAFILGTSSTSLTSSPHSGTSKR 540
Qy 541 RRVTDERELRIPLVGWQRETRIIRNFGRLQGEVAYYAPCGKLRQYPEVIKYLNRNGIM 600
Db 541 RRVTDERELRIPLVGWQRETRIIRNFGRLQGEVAYYAPCGKLRQYPEVIKYLNRNGIM 600
Qy 601 DISRDNFSFSAKIRVGDPIYARDGPQEMQWCLLKEEDVIPRI RAMEGRGRPPNPDROA 660
Db 601 DISRDNFSFSAKIRVGDPIYARDGPQEMQWCLLKEEDVIPRI RAMEGRGRPPNPDROA 660
Qy 661 REESRMRKRGRPPNVGNAEFLDNADAKLRLQAQETARQAQIKLRLKQKQEQARVA 720
Db 661 REESRMRKRGRPPNVGNAEFLDNADAKLRLQAQETARQAQIKLRLKQKQEQARVA 720
Qy 721 KEAKKQQA1MAAEKRRKQKEQIKIMKQKEKIKRIQQIRMEKELRAQQILEAKKKKEBAA 780
Db 721 KEAKKQQA1MAAEKRRKQKEQIKIMKQKEKIKRIQQIRMEKELRAQQILEAKKKKEBAA 780
Qy 781 NAKLLEAKRIKEKEMRRQQAVALKHQBRERRRRQHMLMKAMEARKKAEBEKERLKQEKRD 840
Db 781 NAKLLEAKRIKEKEMRRQQAVALKHQBRERRRRQHMLMKAMEARKKAEBEKERLKQEKRD 840
Qy 841 EKRINKERKLQRLELEMAKELKPNEDMCLADQKPLPELPRIPGLVLSGTSFSDCLMW 900
Db 841 EKRINKERKLQRLELEMAKELKPNEDMCLADQKPLPELPRIPGLVLSGTSFSDCLMW 900
Qy 901 VQFLRNFQKVLGFDVINDVPNLSVLEGLLNI GDSMGVEVDLLVRLLSAAVCDPGLITGY 960
Db 901 VQFLRNFQKVLGFDVINDVPNLSVLEGLLNI GDSMGVEVDLLVRLLSAAVCDPGLITGY 960
Qy 961 KAKTALGHLNLVGNVRNDSVSEILQIFMEAHCGQTELTSKTKFAFOAHTPAQKASVLAF 1020
Db 961 KAKTALGHLNLVGNVRNDSVSEILQIFMEAHCGQTELTSKTKFAFOAHTPAQKASVLAF 1020
Qy 1021 LINELACSKSVVSEIDKNI DVMNLRDKWVVEGKRLRIIHAKTGKTRDTSGSIDIGE 1080
Db 1021 LINELACSKSVVSEIDKNI DVMNLRDKWVVEGKRLRIIHAKTGKTRDTSGSIDIGE 1080
Qy 1081 EQHPLGTPTPGRKRRRKGGSDYDDDDDDDDQDEDEDEDEKEDKQKKTDTICEDED 1140
Db 1081 EQHPLGTPTPGRKRRRKGGSDYDDDDDDDDQDEDEDEDEKEDKQKKTDTICEDED 1140

Qy 1141 EGDQAASVEELEKQIEKLSKQSOYRRKLF DASHSLRSVMFGPDYRRRYRYYWILPRCGGIF 1200
Db 1141 EGDQAASVEELEKQIEKLSKQSOYRRKLF DASHSLRSVMFGPDYRRRYRYYWILPRCGGIF 1200
Qy 1201 VEGMESGEGLEBEIAKEREKKAESVQI KEEMFETSGDSLNCNSNTHDCEQEDLKEKONT 1260
Db 1201 VEGMESGEGLEBEIAKEREKKAESVQI KEEMFETSGDSLNCNSNTHDCEQEDLKEKONT 1260
Qy 1261 NLFLQKPGSFKSLKLEVAKMPPESEVMTPKPNAGANGCTLSYQNSGKHSIGSVQSTAT 1320
Db 1261 NLFLQKPGSFKSLKLEVAKMPPESEVMTPKPNAGANGCTLSYQNSGKHSIGSVQSTAT 1320
Qy 1321 QSNVEKADSNLNTGSSGPGKFYSPLPNDQLLKTLEKNQWFSLLPRTPCDDTSLTHA 1380
Db 1321 QSNVEKADSNLNTGSSGPGKFYSPLPNDQLLKTLEKNQWFSLLPRTPCDDTSLTHA 1380
Qy 1381 DMSTASLVTPQSPSPSPSTPAPLAGSSAQNPVGLNPFALSPLOQVGGSVMGLOFCGW 1440
Db 1381 DMSTASLVTPQSPSPSPSTPAPLAGSSAQNPVGLNPFALSPLOQVGGSVMGLOFCGW 1440
Qy 1441 PTGVVTSNIPPTLSVPISGLSGLSEGNNGSFLTSNVAASSSESPPONEKATSQAQPAV 1500
Db 1441 PTGVVTSNIPPTLSVPISGLSGLSEGNNGSFLTSNVAASSSESPPONEKATSQAQPAV 1500
Qy 1501 EVAKVPDFPSPKPIPEESMOCGMWRIIDPEDLKALLKVLHLRGIREKALQKQIKHLDYIT 1560
Db 1501 EVAKVPDFPSPKPIPEESMOCGMWRIIDPEDLKALLKVLHLRGIREKALQKQIKHLDYIT 1560
Qy 1561 QACLKNKDVA1IELNENEENQVTRDIVENWSVEQAMEMDLSVLOQVEDLERRVASASIQ 1620
Db 1561 QACLKNKDVA1IELNENEENQVTRDIVENWSVEQAMEMDLSVLOQVEDLERRVASASIQ 1620
Qy 1621 VKGWCPEPESEREDLVYFEHKSFTKLCKEHDEGFTGEDES SAHALERKSDNPLDIAVTR 1680
Db 1621 VKGWCPEPESEREDLVYFEHKSFTKLCKEHDEGFTGEDES SAHALERKSDNPLDIAVTR 1680
Qy 1681 LADLERNIERIEEDIAPLRWRRALSEARSAQAVALCICQLOKSI AWEKSIKMYKQCI 1740
Db 1681 LADLERNIERIEEDIAPLRWRRALSEARSAQAVALCICQLOKSI AWEKSIKMYKQCI 1740
Qy 1741 CRKGNBELLLCDGCKGCHTYCHRPKITTI PGDWFPCACIAKASQOTLKI KKLHVKG 1800
Db 1741 CRKGNBELLLCDGCKGCHTYCHRPKITTI PGDWFPCACIAKASQOTLKI KKLHVKG 1800
Qy 1801 KKTNESKKGKVTULTGTDEDEDSASTSSSLKRGNKDLOKRKMEENTSINLSKQESFTSVK 1860
Db 1801 KKTNESKKGKVTULTGTDEDEDSASTSSSLKRGNKDLOKRKMEENTSINLSKQESFTSVK 1860
Qy 1861 KPKRDDS KDLALCSMILTEMETHEDAMPFLPVNLKLVPGYKVKIKKPMDFSTIREKLSS 1920
Db 1861 KPKRDDS KDLALCSMILTEMETHEDAMPFLPVNLKLVPGYKVKIKKPMDFSTIREKLSS 1920
Qy 1921 GOYPNLETFALDVRLVFNDCETFNEDDDSDIGRAGHNMEKYFEKKWTDTFKVS 1972
Db 1921 GOYPNLETFALDVRLVFNDCETFNEDDDSDIGRAGHNMEKYFEKKWTDTFKVS 1972

RESULT 2
US-09-839-479-21
; Sequence 21, Application US/09839479
; Patent No. 6727222
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042002
; CURRENT APPLICATION NUMBER: US/09/839,479
; CURRENT FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JF98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027

; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 21
; LENGTH: 1972
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-839-479-21

Query Match		100.0%	Score 10138;	DB 4;	Length 1972;
Best Local Similarity		100.0%	Pred. No. 0;		
Matches 1972;		Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	MGQTKSTSSGGNRKCNQEQSKNOPLDARVDKIKDKPRKKAMSSSSNSDSGTS	SDTS	60	
Db	1	MGQTKSTSSGGNRKCNQEQSKNOPLDARVDKIKDKPRKKAMSSSSNSDSGTS	SDTS	60	
Qy	61	SEGISSSDDLEDEEEDQISEESDDSDSESAQHKNQVLLHGISDPKADGQA	120		
Db	61	SEGISSSDDLEDEEEDQISEESDDSDSESAQHKNQVLLHGISDPKADGQA	120		
Qy	121	TEKAQEKRIHQPLPLAFESQTHSQSQKQKQVLSQQLPFIQSSQAKESVNHKTS	VIQ	180	
Db	121	TEKAQEKRIHQPLPLAFESQTHSQSQKQKQVLSQQLPFIQSSQAKESVNHKTS	VIQ	180	
Qy	181	STGLVSNVKPLSLVNOAKETMYMLIVPSPDVLKAGNKNTSESSLLTSELRSKRE	QYK	240	
Db	181	STGLVSNVKPLSLVNOAKETMYMLIVPSPDVLKAGNKNTSESSLLTSELRSKRE	QYK	240	
Qy	241	APPSQLKQESSKSLKKVIAALSNPKATSSSPAHPKQTLNHNHPFLTNALLGNHQP	NG	300	
Db	241	APPSQLKQESSKSLKKVIAALSNPKATSSSPAHPKQTLNHNHPFLTNALLGNHQP	NG	300	
Qy	301	VIQSVIQEAPLALTTKTKQKINENIAAASSTPSSPVNLSTSGRRTPGQTPVMP	SAS	360	
Db	301	VIQSVIQEAPLALTTKTKQKINENIAAASSTPSSPVNLSTSGRRTPGQTPVMP	SAS	360	
Qy	361	PILHSQGEKAVSNVNPVKTHSHHPAKSLVEQPRGTDSIPSKKSEDSNEDEE	DE	420	
Db	361	PILHSQGEKAVSNVNPVKTHSHHPAKSLVEQPRGTDSIPSKKSEDSNEDEE	DE	420	
Qy	421	EEDEEDDDSDSDSQSSSDSNTSGSEEDDDKQDESDDTGEKTSMLNKTT	480		
Db	421	EEDEEDDDSDSDSQSSSDSNTSGSEEDDDKQDESDDTGEKTSMLNKTT	480		
Qy	481	SSVKSPSMSLTGCHSTPRNLHIAPGSAAPALCSSESQSPALFGTSSSTLTSS	PHSGTSKR	540	
Db	481	SSVKSPSMSLTGCHSTPRNLHIAPGSAAPALCSSESQSPALFGTSSSTLTSS	PHSGTSKR	540	
Qy	541	RRVTDERELRIPIEYGMORETRIRNFGRLQGEVAYAPCGKLRQYPEVIKYL	SRNGIM	600	
Db	541	RRVTDERELRIPIEYGMORETRIRNFGRLQGEVAYAPCGKLRQYPEVIKYL	SRNGIM	600	
Qy	601	DISRDNFSFSAKIRVGDYEARDGPQENQWCLLKEEDVIPIRAMEGRGRPPNP	DRQRA	660	
Db	601	DISRDNFSFSAKIRVGDYEARDGPQENQWCLLKEEDVIPIRAMEGRGRPPNP	DRQRA	660	
Qy	661	REESRMRRRKGRPNVGNABFLDNADAKLRKLAQAEIARQAQIKLLRKLOKQ	EQARVA	720	
Db	661	REESRMRRRKGRPNVGNABFLDNADAKLRKLAQAEIARQAQIKLLRKLOKQ	EQARVA	720	
Qy	721	KEAKKQQAIAAAEERKQKEQIKTMKQOEKIKR100IRMEKELRAQOILEAKK	KKKEEA	780	
Db	721	KEAKKQQAIAAAEERKQKEQIKTMKQOEKIKR100IRMEKELRAQOILEAKK	KKKEEA	780	
Qy	781	NAKLLEAEKRIKEEMRQQAVLLKHQERRRRRHHMLMKAMARKKAAEKERL	KQEKRD	840	
Db	781	NAKLLEAEKRIKEEMRQQAVLLKHQERRRRRHHMLMKAMARKKAAEKERL	KQEKRD	840	
Qy	841	EKRLNKERKLEQRLELEMAKELKKPNEDMCLADQKPLPRLPGLVLSGST	SDCLMV	900	

Db	841	EKRLNKERKLEQRLELEMAKELKKPNEDMCLADQKPLPRLPGLVLSGSTSDCLMV	900		
Qy	901	VQFLRNFQKVLGFDVNIDVNPNSVLQEGLLNIGDSMGVEVDLLVRLLSAAV	CDPGLITGY	960	
Db	901	VQFLRNFQKVLGFDVNIDVNPNSVLQEGLLNIGDSMGVEVDLLVRLLSAAV	CDPGLITGY	960	
Qy	961	KAKTALGHEHLLNVGNVRDNVSEIIOIIFMEAHCGGTQELTESLKTKAFQAHT	PAQKASVLAF	1020	
Db	961	KAKTALGHEHLLNVGNVRDNVSEIIOIIFMEAHCGGTQELTESLKTKAFQAHT	PAQKASVLAF	1020	
Qy	1021	LINELACSKSVSEIDKNIDYMSNLRDKWVVEKGLKRLRIIHAHKTGKRTSG	SGIDIGE	1080	
Db	1021	LINELACSKSVSEIDKNIDYMSNLRDKWVVEKGLKRLRIIHAHKTGKRTSG	SGIDIGE	1080	
Qy	1081	EQHPLGTPTPGKRRRRKGGSDYDDDDDDDDQDEDEDEDEKEDKQKKT	DI CEDED	1140	
Db	1081	EQHPLGTPTPGKRRRRKGGSDYDDDDDDDDQDEDEDEDEKEDKQKKT	DI CEDED	1140	
Qy	1141	EGDOAASVVEELEKQIEKLSKQSQYRRKLFDAASHLSRVSMFGPDRYRRRY	WILPRCGGIF	1200	
Db	1141	EGDOAASVVEELEKQIEKLSKQSQYRRKLFDAASHLSRVSMFGPDRYRRRY	WILPRCGGIF	1200	
Qy	1201	VEGMESGEGLEEEIAKEREKLLKAESVOIKEEMFETSGDSLNCNTHDCEQ	EDLKEKONT	1260	
Db	1201	VEGMESGEGLEEEIAKEREKLLKAESVOIKEEMFETSGDSLNCNTHDCEQ	EDLKEKONT	1260	
Qy	1261	NLFLOKPGSFSKSLLEVAKMPPSEVMTPKPNAGANGCTLSYQNSGKHS	LGVSQSTAT	1320	
Db	1261	NLFLOKPGSFSKSLLEVAKMPPSEVMTPKPNAGANGCTLSYQNSGKHS	LGVSQSTAT	1320	
Qy	1321	QSNVKEADSNLFWTSGSGPKFYPLPNDQLLTKLTKENRQWFSLLPRT	PCDDTSLTHA	1380	
Db	1321	QSNVKEADSNLFWTSGSGPKFYPLPNDQLLTKLTKENRQWFSLLPRT	PCDDTSLTHA	1380	
Qy	1381	DMSTASLVTPQSQPSKSPSTPAPLGSSAQNQVGLNPPFALSPLQVKG	GVSMGLQFCGW	1440	
Db	1381	DMSTASLVTPQSQPSKSPSTPAPLGSSAQNQVGLNPPFALSPLQVKG	GVSMGLQFCGW	1440	
Qy	1441	PTGVVTGNIPTLSVPSGLSGLGSEGNNGSFLTSNVASSKSESPPV	PONEKATSAQPAV	1500	
Db	1441	PTGVVTGNIPTLSVPSGLSGLGSEGNNGSFLTSNVASSKSESPPV	PONEKATSAQPAV	1500	
Qy	1501	EVAKVPDPPSKPIPEEMQFGWRIIDPEDLALLKVLHLRGIREKALQKQ	IKHLDYIT	1560	
Db	1501	EVAKVPDPPSKPIPEEMQFGWRIIDPEDLALLKVLHLRGIREKALQKQ	IKHLDYIT	1560	
Qy	1561	QACLKNQDVAILIELNENEENQVTRDIYENNSVVEEQAMEMDLSVLQ	VEDLERRVASISQ	1620	
Db	1561	QACLKNQDVAILIELNENEENQVTRDIYENNSVVEEQAMEMDLSVLQ	VEDLERRVASISQ	1620	
Qy	1621	VKGWMCPEPASEREDLVYFEHKSFTKLCHEHDGFTGEDESSAHALER	KSDNPLDIAVTR	1680	
Db	1621	VKGWMCPEPASEREDLVYFEHKSFTKLCHEHDGFTGEDESSAHALER	KSDNPLDIAVTR	1680	
Qy	1681	LADLERNIERRIEDIIAPGLRVWRRALSEARSAQAVALCICQOLQK	SIAMEKSIIMKVYCOI	1740	
Db	1681	LADLERNIERRIEDIIAPGLRVWRRALSEARSAQAVALCICQOLQK	SIAMEKSIIMKVYCOI	1740	
Qy	1741	CRKGDNEELLLLDCGCKGCHTYCHRPKITITIPGDWFCPACIAKASQ	TLIKKLHVKG	1800	
Db	1741	CRKGDNEELLLLDCGCKGCHTYCHRPKITITIPGDWFCPACIAKASQ	TLIKKLHVKG	1800	
Qy	1801	KKTWESKKGKVTUTGTDEDEDSASTSSSLKRGNKDLQKRKWE	NTSINLSKQSF	1860	
Db	1801	KKTWESKKGKVTUTGTDEDEDSASTSSSLKRGNKDLQKRKWE	NTSINLSKQSF	1860	
Qy	1861	KPKRDDSQDLALCSMLITEMETHEDAWPFLPVLNKLVPYKVKV	IKKPMDFSTIREKLSS	1920	
Db	1861	KPKRDDSQDLALCSMLITEMETHEDAWPFLPVLNKLVPYKVKV	IKKPMDFSTIREKLSS	1920	
Qy	1921	GOYFNLETFALDVLVFDNFCETFNEDSDIGRAGHNMRKYFEK	KWTDTFKVS	1972	
Db	1921	GOYFNLETFALDVLVFDNFCETFNEDSDIGRAGHNMRKYFEK	KWTDTFKVS	1972	

```
RESULT 3
US-09-418-710-72
; Sequence 72, Application US/09418710
; Patent No. 6596482
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042001
; CURRENT APPLICATION NUMBER: US/09/418,710
; CURRENT FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 72
; LENGTH: 1969
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-418-710-72

Query Match      98.9%; Score 10022.5; DB 4; Length 1969;
Best Local Similarity 99.4%; Pred. No. 0;
Matches 1962; Conservative 2; Mismatches 4; Indels 5; Gaps 5;

Qy 1 MGQTKSTSSGGNCKNQEQSKNQPLDARVDKIDKPKKAMSSNSDSGTSDDTS 60
Db 1 MGQTKSTSSGGNCKNQEQSKNQPLDARVDKIDKPKKAMSSNSDSGTSDDTS 60

Qy 61 SEGISSSDSDLEDEDEEDQSI ESEDDSDSESAHQKSNQVLLHGISDPKADGKA 120
Db 61 SEGISSSDSDLEDEDEEDQSI ESEDDSDSESAHQKSNQVLLHGISDPKADGKA 120

Qy 121 TEKAQEKRIHOPLAPESQTHSQSQKQOVLSQQLPFIQSQAQKESVNGHTSVIQ 180
Db 121 TEKAQEKRIHOPLAPESQTHSQSQKQOVLSQQLPFIQSQAQKESVNGHTSVIQ 180

Qy 181 STGLSVNKPVLUNQAKKETYMKLI VPSDVLKAGNKNTSEESLLTSELRSKREQYKQ 240
Db 181 STGLSVNKPVLUNQAKKETYMKLI VPSDVLKAGNKNTSEESLLTSELRSKREQYKQ 240

Qy 241 APPSOLKKQESSKSLKVIAALSNPXATSSSPAHPKQTLNNHPNPFILTNALLGNHPNG 300
Db 241 APPSOLKKQESSKSLKVIAALSNPXATSSSPAHPKQTLNNHPNPFILTNALLGNHPNG 300

Qy 301 VIQSVIOEAPLALTTKTMQSKINENTAAASSTPSSPVNLSTSGRTPGNTQTPMPSAS 360
Db 301 VIQSVIOEAPLALTTKTMQSKINENTAAASSTPSSPVNLSTSGRTPGNTQTPMPSAS 360

Qy 361 PILHSQGEKAVSNVNPVKTQHHSHPAKSLVEQFRGTDSDIPSSKDSSEDSNEDEEDDE 420
Db 361 PILHSQGEKAVSNVNPVKTQHHSHPAKSLVEQFRGTDSDIPSSKDSSEDSNEDEEDDE 420

Qy 421 EDEDEDEDDSDSDSQESDSNSDSETEGSEEEEDDDKQDQESDSETEGKTSMLNKTT 480
Db 421 EDEDEE-DEDDSDSDSQESDSNSDSETEGSEEEEDDDKQDQESDSETEGKTSMLNKTT 479

Qy 481 SSVKSPSMSTLGTHTPRNLHTAKAPGAPALCSSESQSPALGHTSSSTLTSSPHSGTSKR 540
Db 480 SS-KSPSMSTLGTHTPRNLHTAKAPGAPALCSSESQSPALGHTSSSTLTSSPHSGTSKR 538

Qy 541 RRVTDRELRPLEYQWQRETRIRNFGRLQGEVAYVAPCGKLRQYPEVIKYLNRNGIM 600
Db 539 RRVTDRELRPLEYQWQRETRIRNFGRLQGEVAYVAPCGKLRQYPEVIKYLNRNGIM 598

Qy 601 DISRDNFSFAKIRVGDFYEARDGQEQMWCLLKEEDVPIPRAMEGRGRGPPNPDQRA 660
Db 599 DISRDNFSFAKIRVGDFYEARDGQEQMWCLLKEEDVPIPRAMEGRGRGPPNPDQRA 658
```

```
Qy 661 REBSRMRRRKGRPPNVGNABFLDNADAKLLRKLQAOEIAQAAQIKLLRKLQKOEQARVA 720
Db 659 REBSRMRRRKGRPPNVGNABFLDNADAKLLRKLQAOE-ARQAAQIKLLRKLQKOEQARVA 717

Qy 721 KEAKKQQAIAAAEBEKRKQEQIKIMKQEQIKRIQIIRMEKELRAAQIILEAKKKKEBAA 780
Db 718 KEAKKQQAIAAAEBEKRKQEQIKIHQEQEKIKRIQIIRMEKELRAAQIILEAKKKKEEAA 777

Qy 781 NAKLLEAEKRIKEEMRRQQAVALLKHOERERRRQHHMLMKAMEARKKAEKERUKQEKRD 840
Db 778 NAKLLEAEKRIKEREMRQQAVALLKROERERRRQHHMLMKAMEARKKAEKERUKQEKRD 837

Qy 841 EKRLNKERKLEQRLELEMAKELKPNEDCLADQKPLPELPRIPGLVLSGTSFSDCLMV 900
Db 838 EKRLNKERKLEQRLELEMAKELKPNEDCLADQKPLPELPRIPGLVLSGTSFSDCLMV 897

Qy 901 VQFLRNFGKVLGFDVNI DVPNLSVLQEG-LLNIGDSMGEVQD LVLRLLSAAVCDPGLITG 959
Db 898 VQFLRNFGKVLGFDVNI DVPNLSVLQEGILLNIGDSMGEVQD LVLRLLSAAVCDPGLITG 957

Qy 960 YKAKTALGEHLLNVGNRDNVSEILQIFMEAHCGQTELTESLTKATQAHTPAQKASVLA 1019
Db 958 YKAKTALGEHLLNVGNRDNVSEILQIFMEAHCGQTELTESLTKAFQAHTPAQK-AVLA 1016

Qy 1020 FLINELACSKSVVSEIDKNIDYMSNLRDRDKWVVEGKLRKLRIIHA KKTGKRDTSGGIDLG 1079
Db 1017 FLINELACSKSVVSEIDKNIDYMSNLRDRKNVVEGKURKURIIHA KKTGKRDTSGGIDLG 1076

Qy 1080 EEQHPGLGTPTPGKRRRKGGSDYDDDDDDSDDDQDQDDEDEDEKEDQKGTDTICEDE 1139
Db 1077 EEQHPGLGTPTPGKRRRKGGSDYDDDDDDSDDDQDQDDEDEDEKEDQKGTDTICEDE 1136

Qy 1140 DEGQAAASVELEKQIEKLKSKQSQYRRKLPDASHSLRSVMFGPDRVRRYRWWILPRCGGI 1199
Db 1137 DEGQAAASVELEKQIEKLKSKQSQYRRKLPDASHSLRSVMFGPDRVRRYRWWILPRCGGI 1196

Qy 1200 FVGEEMSGEGLEETAKEREKLLKKAESVOIKEEMFETSGDSINCSNTHDCHQEKEDLKEKDN 1259
Db 1197 FVGEEMSGEGLEETAKEREKLLKKAESVOIKEEMFETSGDSINCSNTHDCHQEKEDLKEKDN 1256

Qy 1260 TNLFLQKPGSFKSLKLELVAKMPPPESEVMTPKPNAGANGCTLSYQNSGXHSLGVSQSTA 1319
Db 1257 TNLFLQKPGSFKSLKLELVAKMPPPESEVMTPKPNAGANGCTLSYQNSGXHSLGVSQSTA 1316

Qy 1320 TQSNVEKADSNLFPNTGSSGPGKFPYSPLPNDQLLKTUTEKNRQWFSLLPRTPCDDTSLTH 1379
Db 1317 TQSNVEKADSNLFPNTGSSGPGKFPYSPLPNDQLLKTUTEKNRQWFSLLPRTPCDDTSLTH 1376

Qy 1380 ADMSTASLVTPOQPPSKSPSPAPLIGSSAQNPGVGNLPALSPLOVKGVMGLOFCG 1439
Db 1377 ADMSTASLVTPOQPPSKSPSPAPLIGSSAQNPGVGNLPALSPLOVKGVMGLOFCG 1436

Qy 1440 WPTGVVTSNIPFTLSVPSLGSLGSLSENGNSFLTSNVASSKSESPVPQNEKATSAQPA 1499
Db 1437 WPTGVVTSNIPFTLSVPSLGSLGSLSENGNSFLTSNVASSKSESPVPQNEKATSAQPA 1496

Qy 1500 VEKAPVDFPSPKPIPEEMOFGWNRIIDPDLKALLKVLHLRGIREKALQKIQKHLDYI 1559
Db 1497 VEKAPVDFPSPKPIPEEMOFGWNRIIDPDLKALLKVLHLRGIREKALQKIQKHLDYI 1556

Qy 1560 TQACLKXKQVAILTELNEENQVTRDIVENWSVEEQAMWMDLSVLQOVEDLERRVASASL 1619
Db 1557 TQACLKXKQVAILTELNEENQVTRDIVENWSVEEQAMWMDLSVLQOVEDLERRVASASL 1616

Qy 1620 QVKGWMCPEPASBREDLVYFEHKSFTKLCKEHGCEFTGEDESSAHLERKSDNPLDIATV 1679
Db 1617 QVKGWMCPEPASBREDLVYFEHKSFTKLCKEHGCEFTGEDESSAHLERKSDNPLDIATV 1676

Qy 1680 RLADLERNIRRIEEDIAPGLRVRRRALSEARSAQVALCIIQLOKSIAMKSIKMYVCQ 1739
Db 1677 RLADLERNIRRIEEDIAPGLRVRRRALSEARSAQVALCIIQLOKSIAMKSIKMYVCQ 1736
```


QY 1740 ICRKDNBELLLLCDGCDKGCCHYCHRPKITTIPDGDWFCPACIAKASGOTLKIKKJHVK 1799
Db 1737 ICRKGNBELLLLCDGCDKGCCHYCHRPKITTIPDGDWFCPACIAKASGOTLKIKKJHVK 1796
QY 1800 GKKTWESKGGKVVLTGTDEDSASTSSSLKRGKNDJQKRKMBENTSNLSKQSFSTSV 1859
Db 1797 GKKTWESKGGKVVLTGTDEDSASTSSSLKRGKNDJQKRKMBENTSNLSKQSFSTSV 1856
QY 1860 KPKRDDSDDLALCSMILTEMETHEDAWPFLPVNLKLVPGYKVKIKKPMDFSTIREKLS 1919
Db 1857 KPKRDDSDDLALCSMILTEMETHEDAWPFLPVNLKLVPGYKVKIKKPMDFSTIREKLS 1916
QY 1920 SQQYNLETFAVLVDFDNCETFWEDDSDIGRAGHNRKYPKPKWTDTFKVS 1972
Db 1917 SQQYNLETFAVLVDFDNCETFWEDDSDIGRAGHNRKYPKPKWTDTFKVS 1969

RESULT 4

US-09-839-479-71
; Sequence 71, Application US/09839479
; Patent No. 6727222
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042002
; CURRENT APPLICATION NUMBER: US/09/839,479
; CURRENT FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCI/J998/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 71
; LENGTH: 1969
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-839-479-71

Query Match 98.9%; Score 10022.5; DB 4; Length 1969;
Best Local Similarity 99.4%; Pred. No. 0;
Matches 1962; Conservative 2; Mismatches 4; Indels 5; Gaps 5;
QY 1 MGQTKSTSSGGNRKCNQEQSKNQPLDARVDKIKDKPKRKXAMSSSSNSDSGTSSTS 60
Db 1 MGQTKSTSSGGNRKCNQEQSKNQPLDARVDKIKDKPKRKXAMSSSSNSDSGTSSTS 60
QY 61 SEG1SSSDDLDEDEEDDSIESDDSDSEEAHKSNNQVLLHGISDPKADGOKA 120
Db 61 SEG1SSSDDLDEDEEDDSIESDDSDSEEAHKSNNQVLLHGISDPKADGOKA 120
QY 121 TEKAQEKIHQPLAFESQTHSFOSQOQPOVLSQLPFIQSSQAKEESVKNHTSVIQ 180
Db 121 TEKAQEKIHQPLAFESQTHSFOSQOQPOVLSQLPFIQSSQAKEESVKNHTSVIQ 180
QY 181 STGLVSNVKPLSLVNOAKKETYMKLIVPSPDVLKAGNKNTSEESLLTSELRSKREYKQ 240
Db 181 STGLVSNVKPLSLVNOAKKETYMKLIVPSPDVLKAGNKNTSEESLLTSELRSKREYKQ 240
QY 241 APPSOLKQESSKSLKVVTAALSNPKATSSSPAHPKOTLENNHNPFLTALLGNHQNG 300
Db 241 APPSOLKQESSKSLKVVTAALSNPKATSSSPAHPKOTLENNHNPFLTALLGNHQNG 300
QY 301 VIQSVIQEAPLALTTKQKQKINENIAAASSTPSSPVNLSTSGRRTPGNOTPWPAS 360
Db 301 VIQSVIQEAPLALTTKQKQKINENIAAASSTPSSPVNLSTSGRRTPGNOTPWPAS 360
QY 361 PILHSQKKEKAVSNVNPVKTQHHSHHPAKSLVEQFRGTDSDIPSSKDSSEDEEEDDE 420

Db 361 PILHSQKKEKAVSNVNPVKTQHHSHHPAKSLVEQFRGTDSDIPSSKDSSEDEEEDDE 420
QY 421 EDEDEDDDDSDSDSQSSESDSNBSBDTSGSBEEDDDDDKQDESDDTSGEKTSMKLNKTT 480
Db 421 EDEDE-DEDDSDSDSQSSESDSNBSBDTSGSBEEDDDDDKQDESDDTSGEKTSMKLNKTT 479
QY 481 SSVKSPSMSTLGHSTPRNLHIKAPGSAAPALCSSESQSPAFLGTSSSTLTSSPHSGTSGR 540
Db 480 SS-KSPSPSMSTLGHSTPRNLHIKAPGSAAPALCSSESQSPAFLGTSSSTLTSSPHSGTSGR 538
QY 541 RRVTDRELRLPLEYGMORETRIRNFGRLGGEVAYYAPCGKCLROYEVILKYLNRNGIM 600
Db 539 RRVTDRELRLPLEYGMORETRIRNFGRLGGEVAYYAPCGKCLROYEVILKYLNRNGIM 598
QY 601 DISRDNFSFAKIRVGDYFYEARDGPQEMQWCLLKEEDVIPRIRAMEGRGRPPNPDORQA 660
Db 599 DISRDNFSFAKIRVGDYFYEARDGPQEMQWCLLKEEDVIPRIRAMEGRGRPPNPDORQA 658
QY 661 REESRMRRRKGRPNVGNABFLDNADAKLLRKLQAQETARQAQIKLLRKLQKQSOARVA 720
Db 659 REESRMRRRKGRPNVGNABFLDNADAKLLRKLQAQ-ARQAQIKLLRKLQKQSOARVA 717
QY 721 KEAKKQQAIAAAEKKRQKEQIKIMKQOQEKIKRIQOIRMEKELRAQOITLEAKKKKKEBA 780
Db 718 KEAKKQQAIAAAEKKRQKEQIKIMKQOQEKIKRIQOIRMEKELRAQOITLEAKKKKKEBA 777
QY 781 NAKLLEAEKRIKEKEMRRQQAVALLKQHERRRRQHMMLKAMEARKKAEKERLQKQKRD 840
Db 778 NAKLLEAEKRIKEKEMRRQQAVALLKQHERRRRQHMMLKAMEARKKAEKERLQKQKRD 837
QY 841 EKRLNKERKLEQRRLLEMAKELKKPNEDMCLADQKPLPELPIPLGLVLSGSTTFSDCLMV 900
Db 838 EKRLNKERKLEQRRLLEMAKELKKPNEDMCLADQKPLPELPIPLGLVLSGSTTFSDCLMV 897
QY 901 VOFLRNFGKVLGFDVNIIDVNLISVLQEG-LNIGDSMGEVODLLVRLLSAAVCDPGLITG 959
Db 898 VOFLRNFGKVLGFDVNIIDVNLISVLQEGILLNIGDSMGEVODLLVRLLSAAVCDPGLITG 957
QY 960 YKAKTALGEHLNNGVNRDNYSEILQIFMEAHCGOTELTESLTKTKAFQAHTPAQKASVLA 1019
Db 958 YKAKTALGEHLNNGVNRDNYSEILQIFMEAHCGOTELTESLTKTKAFQAHTPAQK-AVLA 1016
QY 1020 FLINELACSKSVVSEIDKNIDYMSNLRDKVWVEGKRLKRIIHAKTGKRDTSGGIDLG 1079
Db 1017 FLINELACSKSVVSEIDKNIDYMSNLRDKVWVEGKRLKRIIHAKTGKRDTSGGIDLG 1076
QY 1080 BEQHPGLTPTGRRKRRKGGSDYDDDDDDSDSDQDDEDEDEDEDEDEDEDEDEDEDEDE 1139
Db 1077 BEQHPGLTPTGRRKRRKGGSDYDDDDDDSDSDQDDEDEDEDEDEDEDEDEDEDEDEDE 1136
QY 1140 DEGOAASVELEKQIEKLSKQSQYRRKLPDASHLSRSMFDPDRYRRRYWILPRCCGI 1199
Db 1137 DEGOAASVELEKQIEKLSKQSQYRRKLPDASHLSRSMFDPDRYRRRYWILPRCCGI 1196
QY 1200 FVEGMESGEGLEEIAKERELKKAESVQIKKEMFETSGDSLNCNSNTRDCEQKEDLKEKDN 1259
Db 1197 FVEGMESGEGLEEIAKERELKKAESVQIKKEMFETSGDSLNCNSNTRDCEQKEDLKEKDN 1256
QY 1260 TNLFLOKPGSFSLSKLLEVAKMPPESEVMTPKPNAGANGCTLSYQNSGKSLGSLVQSTA 1319
Db 1257 TNLFLOKPGSFSLSKLLEVAKMPPESEVMTPKPNAGANGCTLSYQNSGKSLGSLVQSTA 1316
QY 1320 TQSNVEKADSNLNFNTSGSGPKFVPLPNQDLKLTTEKRNQFSLPRTPCDDTSLTH 1379
Db 1317 TQSNVEKADSNLNFNTSGSGPKFVPLPNQDLKLTTEKRNQFSLPRTPCDDTSLTH 1376
QY 1380 ADMSTASLVTQSQPPSPKSPSTPAPLGSSAQNPVGLNPFALSLPLQVKGVSMMGLQFCG 1439
Db 1377 ADMSTASLVTQSQPPSPKSPSTPAPLGSSAQNPVGLNPFALSLPLQVKGVSMMGLQFCG 1436
QY 1440 WPTGVVTSNIPTLTVSPSLGSLGSEGNNSFLTSTNVASSKSPVPQNEKATSAQAPAA 1499
Db 1437 WPTGVVTSNIPTLTVSPSLGSLGSEGNNSFLTSTNVASSKSPVPQNEKATSAQAPAA 1496

```
Qy 1500 VEVAKPVDFFSPKPIPEMQFGWRIIDPEDLKALLKVLHLRGIREKALQKIQKHLDIYI 1559
Dy 1497 VEVAKPVDFFSPKPIPEMQFGWRIIDPEDLKALLKVLHLRGIREKALQKIQKHLDIYI 1556
Qy 1560 TOACLKNDKVAIIELNENEENQVTRDIVENWNSVEEQAMEMDLSVLQOVEDLERRVASASL 1619
Dy 1557 TOACLKNDKVAIIELNENEENQVTRDIVENWNSVEEQAMEMDLSVLQOVEDLERRVASASL 1616
Qy 1620 QVKGWMCPEPASEREDLVYFHKSFYKLCCKHDEGTEGDESSAHALERKSDNPLDIATV 1679
Dy 1617 QVKGWMCPEPASEREDLVYFHKSFYKLCCKHDEGTEGDESSAHALERKSDNPLDIATV 1676
Qy 1680 RLADLERNIERIEDIAPGLRVWRRALSEARSAAQVALCIQOQKSIWAEKSIWKVYCYQ 1739
Dy 1677 RLADLERNIERIEDIAPGLRVWRRALSEARSAAQVALCIQOQKSIWAEKSIWKVYCYQ 1736
Qy 1740 ICRKGDNEELLCDGCGDKGCHTYCHRPKITTIPDGWDFCPACIAKASGGQTLKIKLHVK 1799
Dy 1737 ICRKGDNEELLCDGCGDKGCHTYCHRPKITTIPDGWDFCPACIAKASGGQTLKIKLHVK 1796
Qy 1800 GKKTNEKKGKVTITGDTEDDSASTSSSLKRGKDKOLQKCKMEENTSIINLSKQESFTSV 1859
Dy 1797 GKKTNEKKGKVTITGDTEDDSASTSSSLKRGKDKOLQKCKMEENTSIINLSKQESFTSV 1856
Qy 1860 KKPXRDDSKOLALCSMILTEMETHEDAWPFLLPVNLKLVPGYKVKIKPMDFTSIREKLS 1919
Dy 1857 KKPXRDDSKOLALCSMILTEMETHEDAWPFLLPVNLKLVPGYKVKIKPMDFTSIREKLS 1916
Qy 1920 SGQYPNLETFALDVLVFDNCETPNEDSDSDIGRAGHNNRKYFEKKWTDTFKVS 1972
Dy 1917 SGQYPNLETFALDVLVFDNCETPNEDSDSDIGRAGHNNRKYFEKKWTDTFKVS 1969
```

RESULT 5

```
US-09-418-710-71
; Sequence 71, Application US/09418710
; Patent No. 6596482
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042001
; CURRENT APPLICATION NUMBER: US/09/418,710
; CURRENT FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 71
; LENGTH: 1876
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-418-710-71
```

Query Match 24.6%; Score 2493.5; DB 4; Length 1876;

Best Local Similarity 32.1%; Pred. No. 2.5e-168;

Matches 675; Conservative 343; Mismatches 674; Indels 414; Gaps 66;

```
Qy 43 MESSNSDSD-----SGTS-SDTSSSEGISSSSDSDDEDEEEDQSIIESED 88
Dy 1 MEMEANEANDHFNFTGLPPAPASGLKPSGSGGLTYNGS---PHNFPQOGKSL--NGD 55
Qy 89 DSDSSESAQKSNQVLLHGISDPKADGQKATEKAQEKRIHQPLPLAFE-----SQTHS 143
Dy 56 VVNGLSVTSHTTSGIL-----NSAPHSSTSHLHP-SVAYDCLWNYSQVPS 103
Qy 144 FQ--SQKQPVLVSO---QLPFIQSSQAKESVNKHTSVIQ-----STGLVSNVNP 190
Dy 104 ANPGSNLKDPPLLSQFGGQVFLNGILGSGRQSPSPSHNTNLRAGSKQFWANGTHS---P 160
```

```
Qy 191 LSLVNOAKKETYMKL-----IVSPDVLKAGNKNTSBESSLLTSELRSKREQ-- 237
Dy 161 MGL-NFDSQBLYSFPDQNFEEVCSGHPDEAAEKEMTSVVAENGTLGLVCSLELEEQPE 219
Qy 238 -----YKQAPPSQKQKQESSKSKVKVIAALSNPKATSSSPAHPKOTLENNHNPFLNAL 292
Dy 220 LKMGYNGSVF-----SVESLHQEVSVLV-PDPTVSCLDLP----- 254
Qy 293 LGNHQPNGVITQSVIOEAPLALITTKMQSKINENIAAASSTPPSSPNVLSGSRRTQNG 352
Dy 255 --SHLPD-----QLEDTP-----LSEDSLEPFNSLAPPEPVSGGLYGI DD 292
Qy 353 TPVMPSPASPIILHSGGKEKAVSNVNNVPKTV-----QHSHHPAKSLVEQFRGTDSDIPESKD 407
Dy 293 TELM-----GAEDKLPLEDSFVISALDCPSLNNATAFSLADDSQTSIIFASPT 342
Qy 408 SEDSNDEDEDEDEDEDEDEDDSDSQSESNSNESDTEGSEEEEDDDKQDQESDSDT 467
Dy 343 SPPVLGSEVLQDNSFDLNNGSDAEQEBMETQSSDFPPLTOPA-----PDQSSTIQ 393
Qy 468 EGEKTSMKLNKTTT-----SFKSPMS-----SVKSPMS----- 489
Dy 394 LHPATSPAVSPPTTSPAVSLVSPAAASPEISPEVCPAAASTTVVSPAVFVSPASSAVLPAV 453
Qy 490 ----LTHGSTPRNLHIKAPGSA-PAALCSESQSPAFLTSSST-----LTSSPHSG 536
Dy 454 SLEVPLTASVTSPKASPVTSFAAAPTASPAKNDVSSFLTTADVETITGEGLTAS--GSG 512
Qy 537 TSKRRRVTRDELRIPLRYGWQRETRIRNFGRLQGEVAYYAPCGKKLRQYVEVIKYLRS 596
Dy 513 DVMRRRIATPEEVRLPLQHGWRREVRIKKGSHRWQGETWYYPGCGKRMKQFPEVIKYLRS 572
Qy 597 NGIMDISRDNFSAKIRVCDGFYEARDGPQEMQWCLLKKEEDVIPRIAMGRRGRPPNDP 656
Dy 573 NLVHSVREHFSPSPMPVGDFFPEEDTPEGLQWVQSABEIPSRIOAITGKGRPPNTE 632
Qy 657 RORABESRMRRKGRPPNVGNAEFLLDNADAKLLRKLQAOEIAQAAQIKLLRKLQKOEQ 716
Dy 633 KATKEVPKVRGRGRPPKVIITELNKTNRPLKLEAQE-----TLNEDK 680
Qy 717 ARVAKEAKQQAIAAAEERKQKEQIKIMQKQEKIKR IQIRMEKELRAQAQIIEAKKKK 776
Dy 681 AKIAKSKKMR-----QKVGEGCLTITIQOARNK-----PKQETKSLKHKEAKKKSK 728
Qy 777 BEAANAKLLEAKRIKKEKVRQOAVLLKHQERERRRQHMLMKAMEARKKAEKERLKQ 836
Dy 729 AEKERGK--TKQEKLEK-----VKREKKEK-----VKKEEVEVTTAKP 765
Qy 837 EKRDEKRLNKERKLEQRRLSEMAKELKPNEDMCLADQKPLPELPRIPLGLVLSGSTFSD 896
Dy 766 ACKADKTLATORLEERKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQ 825
Qy 897 CLMVVQFLRNFGKVLGFDVNDIVPNLSVLQEGLLNIGDSMGVEQDQLIVRLLSAACVDPGL 956
Dy 826 CLTIVEFLHSPGKVLGFDPAKDVPSLQEGLLCQGDLSGVEQDQLIVRLLKAAALHDPGF 885
Qy 957 ITGYKAKTALGEHLLNVGNRDNVSEILOIFMAHCCQOTELTSELTKTAQAHTPAQKAS 1016
Dy 886 PSYCQSKKILGEKVESEIPLTRDNVSEILRCFLMAYGVEPALCDRLRTPQFQAQPPQKAA 945
Qy 1017 VLAFLINELACSKSVSEIDKNIDYMSNLRDRKVVVEGKRLKRLRIHAKTKGRDRTSGGI 1076
Dy 946 VLAFFVHELNGSTLIINEIDKTLIESMSYKKNKWIVEGRRLRLTLVLAKTGRSE----V 1001
Qy 1077 DLGEEHQPLGTPTKRRRKRGSDSDYDDDDDDDDQDDEDEDEDEDEDEDEDEDEDEDEDEDE 1136
Dy 1002 EMGRPEBCLG-----RRSSRIME-----ETSGMEEEEEEESIAAVPGRRK-- 1043
Qy 1137 EDEDEGQAAASVELEKQIEKLSKQSQYRKLFDASHSILRSVMFGDRVRRRRVWILPRC 1196
Dy 1044 RDGEVDATASSIPELERQIEKLSKQLFRKRLHSSQMLRAVSLGQDRVRRRRVWILPYL 1103
```

```
QY 1197 GGFVFGMESGRLBE-TAKEREKLLKAESVOIKEMP-----ETSGDSLNCSTNTHCRQK 1251
Db 1104 AGIFVEGTEGNLVPEVIVKETTSDSKVAHAASLNPALFPMKLMELAGSNTTASSPARARSR 1163
QY 1252 EDLKEKDNTNLFLQKPGFSKLSKLE---VAKMPPESEVWTP-KPNAGANGCTLSY-- 1304
Db 1164 -PLKTKPG---FMQ--PREFKSPVRQDSEFQPOAQLOPEAQLHVPAQFQLOLQLOSHKG 1218
QY 1305 ---QNSGKHSGSVOSTATQSN-----VEKADSNLNFN---TGSSGPKFYSPLPND--- 1350
Db 1219 FLEQSGSPLSLQSQHDSQSASFLSWLSQTOHSSLLSSSVLTPDSSPGKL-DPAPSQPPE 1277
QY 1351 ----QLLXLTLEKNQWESLLPRTCCDTSLTADMTASLVTQSQPPSKSPSTPAPL 1406
Db 1278 EPEPDEASSPDLQAFWNISAOFCN-----AAPTPLVASEDQFTSPQOL 1325
QY 1407 GSS---AQNVPGLNPFALPQVKGGVMMGL---QFCGWPTGVVTSNIPFTLSVPSLSG 1461
Db 1326 ASSKPMNPSAANP---CSFVQF-SSTPLAGLAPKRRAGDPG-----EMPQSPFG 1371
QY 1462 LGLSEGNNGSFLTSNVASKSSESPPVQNEKATSAQPAAVEVAKPVDPPSPKPIPEMOPG 1521
Db 1372 LQOPKRRG-----PPSKF---FKQMEQRLVTQLTA-----QPVPPEMCSG 1409
QY 1522 WRRIIDPEDLKALLVHLRGIREKALOKIQKHLDYITQACLKNKQVAIIELNENBQ 1581
Db 1410 WWWIPDPMDLAKALHPRGIREKALHKLNRDFLQEVCLR---PSADPIPEPROLP 1466
QY 1582 VTRDIVEWMSVEQAMWDLVLQVEDLEREASASLQVGMWCPESPASEREDLAVPEH 1641
Db 1467 AFQEGIMWSPEKETYETDLAVLQWVELEQKRVMSDQIRGWTCPSPDSTREDLACEH 1526
QY 1642 KSFTXLCHEHGEFTGEDESSAHLERKSDNPLDIATVRLADLERNIERR----- 1691
Db 1527 LSDS-----QEDITWGPGRGL-AQPKTNPDLAVMLAALAEQNVKRYLREPLWPTH 1581
QY 1692 -----IEDIAPGLURVRRALSEARSAAQVACLCIOLOKSTAW 1729
Db 1582 EWLEKALLSTNGAPEGTTTSEIETPRIMRQTLQRCRSAARHVCCLGHLERSIAW 1641
QY 1730 EKSIMKVCQCRKDNELLILCDGCKGCHTYCHRPKIITIPDGNFPCACIAKASGQ 1789
Db 1642 EKSVMKVTCLVCRKDNDEFLLLCDGCRGCHTYCHRPKMEAVPEGDWFCVTCLAQ---- 1697
QY 1790 TLKIKLHVKGKTYNE---SKKGKVTILTGTODEDSASTSSLRKGNKDLQKRKEENT 1846
Db 1698 -----QVEGFTQKPFPGKQK-----RKSGYSLNFSEGDGRRRVLKGRESAA 1744
QY 1847 SINLSKQSFSTSVKXP--KRDSKDLALCSMILTEMETHEDAWPFLLPVNLKLVPGYKXV 1904
Db 1745 GPRYSEERLSPSKRRPLSRNHHSDLTFCEIILMEMESHDAWPFLEPVNPLVSGYRRI 1804
QY 1905 IKKPNDFSTIREKLSSGQYPNLETALDVLRLVDFNCETFNEDDSIDIGRACHNMRKYFEKK 1964
Db 1805 IKPNMDFSTRERLRGGYTSSEEPADALLVDFNCQTFNEDDSVKGAGHIMRRFFESR 1864
QY 1965 WTDTEK 1970
Db 1865 WEEFYQ 1870
```

RESULT 6

```
US-09-839-479-70
; Sequence 70, Application US/09839479
; Patent No. 6727222
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042002
; CURRENT APPLICATION NUMBER: US/09/839,479
; CURRENT FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 09/418,710
; PRIOR FILING DATE: 1999-10-15
```

```
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 70
; LENGTH: 1876
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-839-479-70
```

Query Match 24.6%; Score 2493.5; DB 4; Length 1876;

Best Local Similarity 32.1%; Pred. No. 2.5e-168;

Matches 675; Conservative 343; Mismatches 674; Indels 414; Gaps 66;

```
QY 43 MESSNSDS-----SGTS-SDTSSEGISSDSDDEDEDEEEDSIESD 88
Db 1 MEMEANEANDHFNTGLPPAPAAAGLKPSPSGEGLYTNCS---PHNFPQOKSL--NGD 55
QY 89 DSDSSESAQHSNNQVLLHGISDPKADGQKATEKAQEKRIHQPLPLAFE-----SOTHS 143
Db 56 VVINGLSTVSHTTTSGIL-----NGAPSSSTSHLHP-SVAYDCLWNYSQYPS 103
QY 144 FQ--SQOKQPOVLSO---QLPFIQSSQAKESVNHKTSVIQ-----STGLVSNVKP 190
Db 104 ANPGSNLKDPLLSQFSGQYPLNGILGSGRQSPSPSHNTNLRAGSKFWANGTHS---P 160
QY 191 LSLVNAQAKFTYMKL-----IVPSDDVLKAGNKNTSSESSLTSELRKREO-- 237
Db 161 MGL-NFDSQELYDFPDQNFBEVCSGIHPDEAAEKEMTSVVAENGTGLVCSLEBEEQPE 219
QY 238 ----YKQAFPSQLKQESSKLVIAALSNPKATSSPAHPKQTLNHNHNPFLTNAL 292
Db 220 LKMGYNGSVP-----SVESLHQEVSLV-PDPTVCLDDP----- 254
QY 293 LGNHOPNGVIOQVIOEAPLALTTTKMQSKINENIAAASSTPFSFVNLSSTSGRRTPGQ 352
Db 255 --SHLPD-----QLEDTPI-----LSESLPEFNSLAPFVSGGLYDID 292
QY 353 TPVMPSPAILHSGQEKAKAVSNVNPVKT-----QHSHPAKSLVEQFRGTDSIPSSKD 407
Db 293 TELM-----GAEDKLPLEDSPVISALDCPSLNNATAFSLADDSTSTSFASPT 342
QY 408 SEDSNEDEDEDEDEDEDEDDSDSQSSEDSNSESDETEGSEEDDDDDKQDSDSDT 467
Db 343 SPPVLGESVLQDNGSFDLNGSDAEQEMETOSSDPFSLTOPA-----PDQSSTIQ 393
QY 468 EGKTSMKLNTTS-----SVKSPSMS----- 489
Db 394 LHPATSPAVSTTTPAVSLVSPAASPEISPEVCPAASTVVSVPVSVSPASSAVLPAV 453
QY 490 ----LTGHSTPRNLHTAKAFCSA--PAALCESQSPAFILGTSSST-----LTSPPHSG 536
Db 454 SLEVLPTASVTSPKASPVTSAAAAPFTASPAKDVSSFLETTADVEEITGEGLTAS--GSG 512
QY 537 TSKRRRVTDRELRIPLEYGHWORETRIRNFRGLOGEVAYYAPCKKRLQRYPEVILKYLRS 596
Db 513 DVMRRIATPEEVALPLQHGWRREVRIRKKGSHRWQGETWYTGPCGRMKQFPFVILKYLRS 572
QY 597 NGIMDISDNFSFSAKIRVGDYFEGDQPOEMONCLLKEEDVIPIRAMEGRGRPPNPD 656
Db 573 NLVHSVRKEHFSFSPMPVPGDFFERDTPPEGLQWQLSABEIPSRQAITGRGRPNRTE 632
QY 657 RQAREESRMRKRGRPNVNGNABFLDNADAKLRLQAEIARQAQIKLRLKQKQKQ 716
Db 633 KAKTYEVPKVRGRGRPEPKVITELLNKTNDNRPLKLEAQE-----TLNEEDK 680
QY 717 ARVAKEAKQQAQIMAAEERKQKQEQIKIMQOEKIKRIQIRMEKELRAQOILEAKKKK 776
Db 681 AKIAKSKKKMR-----QKVQGECLTTIQGARNK-----FKQETKSLKHKEAKKSK 728
```


Db 702 DACMELRLSNPSLVKLSSTSVYDLTPGKMKILHALCGKLL---TLVSTRDFIEDYVDI 758
Qy 1045 LRRDKWVV-----EGKLR-----KLRIHAKTKGRTDTS 1075
Db 759 LRQAQEFRELKAEQHRKERBEAAARIRKREKLEKEQKMKKEDEQKNSAD 818
Qy 1076 IDLGEQHPGLGTPGKRKRKGGSDYDDDDDDDDQDDEDEDEDE---KEDQK 1131
Db 819 ISIGEE-----REDFTSIESKDTQEKELDQDMFTDEDDPGSHKGRGK 865
Qy 1132 KTD-----ICEDEGQAAASVELEKOIKLSKQSQYRRKLFDA 1172
Db 866 RGQNGFKEFTROEQINCVTRELLTADEE---ALKQEHQKELKEKIQS-----A 914
Qy 1173 SHLSRVNFGDRYRRYRRIIPRCGGIFVEGMESEGLEEIAKERKLLKXAESVQIKEM 1232
Db 915 IACTNIFPLGRDRYRRIIPFSPGLFIE--EDYSGLTE-----952
Qy 1233 FETSGDSLNCNTDHCQEKEDLKEKDNLTNLFQKPGFSKLSKLEAVAKMPPESEVMPK 1292
Db 953 -----DMLLPRSSFQ-----NVQSDPQVSTKTGE 979
Qy 1293 PNAGANGCTLSYQNGKHSLSVOSTQSNVEKADSNLNTGSSGPKFYPLPND-- 1350
Db 980 P-----LMSEST-SNIDQG-----PRDHS 997
Qy 1351 -OLLKTLTEKRWQFSLPRPCDDTSLTHADMSTASLVTQSPQPSKSPSPTAPLGSS 1409
Db 998 VOLPKPVHKPNR-W-----CFYSSCEQLDQLEAL-----1026
Qy 1410 AQNPVGLNPFALLPLQVKGVSMMGLQFCGWPTGVVTSNIPFTLSVPSGLGLSEG 1469
Db 1027 -----NSRGHR 1032
Qy 1470 NSFILTSNVASSKSPVPQNEKATSAQAAPAAVEVAK--PVDFPSPKPIPEEMQFGWRIID 1527
Db 1033 ESALKETLLQEKSR-----ICAQLARFSEKHFSDKFPQDPKSKPTYSKGRSSNAYD 1083
Qy 1528 PEDLKALLKVLHLR-----GIREKALQ-----KQIQKHLDYITQACLKNKDVAILTN 1575
Db 1084 PSQWCA-EKQLELRDLFLDIEDRIYQGTIGATKVTDH---IWRSALESGRYELLS-E 1138
Qy 1576 ENEENQVTRDVTNVMNSVEQAMEMDLSVLQOVEDLERRVASASLQVKGWMCPEPASERED 1635
Db 1139 ENKENGIIKTV--NEDVZEMIDEQTKVIVKDRLLGKITETPSTVSTWASTPQSVS-----1192
Qy 1636 LVYFEHKSFTKLCKEHGDEFTGEDESSAHALERKSDNPLDIAVTRLADLERNIERRIEED 1695
Db 1193 -----SVVHYL-----AVALFOIEQGIERRFLK- 1215
Qy 1696 IAP-----GLRVWRRLSEARSAAQVACIOLOKSTAWESIMKVYQOICRK 1743
Db 1216 -APLDASDSGRSYKTVLDRWRRESLSSASLQVFLHLSTLDRSVTWSKSIINARCKICRK 1274
Qy 1744 GDNELLLLDCGCKGCHTYCHRPKITIPDGMFCPCACIAKASQOTLKIKLHVKGKKT 1803
Db 1275 KGDAENMVVLCDCDGHHTYCVRPKLKTVPEDWFCPCRPQ-----RCRRLSFRQPS 1329
Qy 1804 NESKK-----GKKVTLTGDTE-----DED-----1822
Db 1330 LESDEVEDSMGGEDEVDGDEEGSQSEEEVEVEQDEDDSQSEEEVSLPKRGRPOVRLP 1389
Qy 1823 -----SASTSSSL-----1830
Db 1390 VKTRGKLSFSFSRQOQEPGRYPSRSQOSTPKTTVSSKTRSLRKINSAPPTETKSURI 1449
Qy 1831 -----KRGKND-----LQKRKMEENTSI 1848
Db 1450 ASRSTRSHGPLQADVVELLSPRKRRGRKSANNTPNSPNFNRVIAKTSQEQSRSV 1509
Qy 1849 N-----LSQSEFTSVKPKRDSK-----DLALCSMILTEMETHEDA 1886
Db 1510 NIASKLSLQSESKRRCRKRPSPSVTLGRRSSGRQGVVHLSAPQLVVELVRHDDS 1569

Qy 1887 WPFLLPVNLKLVPGYKVKVKKPMDFTSTIREKLSGSGOYNLETFALDVRVLFDNCETFNED 1946
Db 1570 WPFLLKLSKVQVPDYDIKKPIALNIRKVNKCEYLASEFIDDIELMESNCFEYNPR 1629
Qy 1947 DSDIGRAGHNMRYF 1961
Db 1630 NTSEAKAGTRLQAFF 1644
RESULT 12
US-09-839-479-1
; Sequence 1, Application US/09839479
; Patent No. 6727222
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042002
; CURRENT APPLICATION NUMBER: US/09/839,479
; CURRENT FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1674
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-839-479-1
Query Match 7.2%; Score 726.5; DB 4; Length 1674;
Best Local Similarity 21.5%; Pred. No. 1.8e-42;
Matches 338; Conservative 186; Mismatches 424; Indels 627; Gaps 53;
Qy 664 SMRRRRKGRPPNVGNABFLDNADAKLLRLKLOAQAFIAQAAQIKLLRLKLOQEQARVAKEA 723
Db 420 SPANRRGRPP-----KRIHISQEDNVA--- 442
Qy 724 KQQAIAAAEAKKQKQKIKIMQOEKIKRITQOIRMEKELAAQOILEAKKKKEEANA 783
Db 443 --NKQTASVRSKATKDKRLKQEE---MKSIAFE-----KAKLKKREKADA- 484
Qy 784 LLEAKRIKEMRRRQOAVLLKHQERRRRRHHMLKAMEARKKAEKERLQEKRDDEKR 843
Db 485 -LEAKK--KEKE-----DKEKKRELKKIVEEERLKKKEKERLKVREKE-- 527
Qy 844 LNKERLEQRRLELEMAKELKKPNEDMCLADQKPLPELPRIIPGLVLSGSTFSDCLMVVQF 903
Db 528 --REKLREKRYVEYLKQWSPKPREMCECDLKLPE-PTPVKTRLPEIIPGDMALMVLEF 584
Qy 904 LRNFQVLGPDVNDVFN---LSVLQEGLLNIGSMGEVQDILLVRLLSA---AVCDPLGI 957
Db 585 LNAFGEI--FDLQDEFPGVTLEVEALVG-NDSEGPLCELLFFFLTAIQAIAESEE 641
Qy 958 TGYKAKT---ALGBHLLNVGNVRDNVSEILQIF-----987
Db 642 VAKQLTDADTKGCSLSKSLDSDCTLSEILRLHLASGADVTSANAKRYQKRGFADTD 701
Qy 988 ---NEAHCGQTELTESLTKAFOAHTPAQKASVLAFLINELACSKSVSVSEIDKNIDYMSN 1044
Db 702 DACMELRLSNPSLVKLSSTSVYDLTPGKMKILHALCGKLL---TLVSTRDFIEDYVDI 758
Qy 1045 LRRDKWVV-----EGKLR-----KLRIHAKTKGRTDTS 1075
Db 759 LRQAQEFRELKAEQHRKERBEAAARIRKREKLEKEQKMKKEDEQKNSAD 818
Qy 1076 IDLGEQHPGLGTPGKRKRKGGSDYDDDDDDDDQDDEDEDEDE---KEDQK 1131


```
Db 914 IACTNIFPLGRDRMYRRYWIFFSLPGLFIE--EDYSGLTEDHLLPRPSSFNQNVQSDPQ 971
Qy 1233 PET-SGDSLNCSTHDCQKEDLKEKDNINFLQKPGSFKSLKLELVAKMPPSEVWTP 1291
Db 972 VSTKTGEPLMBESTSNIDQG-----PRHVSQVLP 1000
Qy 1292 KPNAGANGCTLSYQNSGKHSLGSVQSTATQSNVEKADSNLNTFNTGSSGPGK--FYSPLPN 1349
Db 1001 KP-----VHK-----PNRWCFYSCEQ 1017
Qy 1350 -DQLKTLTKNRQWFSLLPTPCDDTSLTHADMSTAS-----LVTPSQSPFSK----- 1397
Db 1018 LDQLTEALNSGRHR--ESALKETLLQKSRICAQLARFSEKPFHFSKRPQSDSKPTYSRGR 1076
Qy 1398 -SPSTPAPLSSAQNPLGAPFALSPLQVKGVSMMGLQFCGWPVTGVVTSNIPFTLSVP 1456
Db 1077 SSNAYDPSQCAEKQLELRDLFL-----DIEDRIYQG 1110
Qy 1457 SLGSLGLSEGNSTNSVASSKSPVQNEKATSAQPAAVEVAKPVPFPKPIPE 1516
Db 1111 TLG-----AKVTRHWR----- 1125
Qy 1517 EMQFGWRIIDPEDLKALLKVLHLRGIREKALQKIQKHLDYITQACLKNDKVAIIELNE 1576
Db 1126 ALESGRYELLSEEN-----KENGIIKTV-----NEDVEEMEIDE 1159
Qy 1577 NEENQVTRDIVENSVESQAMEMDLVLTQVEDLERRVASASLQVGMCEPESERDL 1636
Db 1160 QTKVIVKDRLLGKITETSTVSTNASTPQSVSVVHYLAMALFOIE----- 1205
Qy 1637 VYFEHKSPTKLCKEHGDEFTGEDESSAHALERK--SDNPLDIAVTRLADLERNIERRED 1695
Db 1206 -----QCLERRFLKAPLD-----ASDSGRSYKTVLDR- 1232
Qy 1696 IAPGLRVRRALSEARSAQVLCIQLOKQIAWEKSIKMYCQICRKGDNELLLCDG 1755
Db 1233 -----WRESLLSSASLSQVFLHLSTLDRSVIWSKSIILNARCKICRKGDAENNVLCG 1285
Qy 1756 CDKGCHTYCHRPKITITIPDGWFCPACTAKASGQTLKIKLHVKGKTKNESKK----- 1808
Db 1286 CDRGHHTYCVRPKLIKIVPEGDFWFCPCRPKQ-----RCRRLSFRQPSLESDDEDVDSMG 1340
Qy 1809 GKVYTLTGDT-----DED----- 1822
Db 1341 GEDDEVGDREGEQSEEEVEVEQEDDSDQSEEVSLPKRGPQVRLPVKTRGKLSSESFS 1400
Qy 1823 -----SASTSSSL----- 1830
Db 1401 SRGOQOEPGRYPSRSQSQSTPTKTVSSKTGRSLRKINSAPPTETKSLRIASRSTRSHGPL 1460
Qy 1831 -----KRGKND-----LOKRWENTSN-----LSKQBSF 1856
Db 1461 QADVVELLSRPRKRGKRSKANNTPNSPNFNFRTVATKSSQSRSVNIASKLSQSE 1520
Qy 1857 TSVKPKRDKSK-----DALCSMLTSMETHEDAWPFLLPVNLKLV 1898
Db 1521 SKRCRKQKQSPSPVTLGRSSGQGVHLSAFEQULVVELVRHDDSWPFLKLVSKIQV 1580
Qy 1899 FGYKVIKKMPDFSTIREKLSGQYPNLETFALDVRLVDFNCETFNEDDSDIGRAGHNR 1958
Db 1591 PDYDIIKKPIALNIIIREKVNKCEYKLASEFIDDIELMFSCFENPNTSEAKAGTRLQ 1640
Qy 1959 KYF 1961
Db 1641 AFF 1643
```

RESULT 14

```
US-09-839-479-69
; Sequence 69, Application US/09839479
; Patent No. 6727222
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
```

```
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042002
; CURRENT APPLICATION NUMBER: US/09/839,479
; CURRENT FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 69
; LENGTH: 1673
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-839-479-69

Query Match 7.0%; Score 709; DB 4; Length 1673;
Best Local Similarity 21.0%; Pred. No. 3.2e-41;
Matches 329; Conservative 204; Mismatches 426; Indels 604; Gaps 51;

Qy 664 SRMRRRGRPRPNVGNASFPLDNADAKLRKLQAQBIARQAQAIKLRKLQKQEQARVAKEA 723
Db 420 SPANRRGRPP-----KRIHISQEDNVA--- 442
Qy 724 KKOQAIMAAEERKQKQSIKIMKQEKIKRQIQIRMEKELRAOQILEAKKKKEEANAOK 783
Db 443 --NQTLASYSRKATSRDKLLKQEE---MKSLEAF-----KAKLKREKADA- 484
Qy 784 LLEAEKRIKEKEMRRQQAVALLKQERERRRRQHMMMLKAMEARKKAEKERLQKQREKDR 843
Db 485 -LEAKK--KEKE-----DKEGKEELKKIVEEELKKKEKEKELKVEREKE-- 527
Qy 844 LNKERKLEORLELEMAKELKPNEDCLADQKLPPLPRIPGLVLGSGTSTSDCLMVVQF 903
Db 528 --REKLREKKYVEY--KQMSKPREDMBCDDLKELPE-PTPVKTRLPPEIFGDALMWLEF 583
Qy 904 LRNFQKVLGFDVNTDVFN---LSVLQGLNLIGNDSMGEVQDILLVRLLSA---AVCDPGLI 957
Db 584 LNAFGEL--FDLQDEFPDGVTLVELEALVG-NDSEGPLCELLFFFLTAIFQAIAEBEE 640
Qy 958 TGYAKT---ALGSHILNVGNRDNVSEILQIF----- 987
Db 641 VAKQLTDADTKGCSLKSLLDLSCTLSLSEILRLHLTLASGADVTSANAKYVKRGGFDATD 700
Qy 988 ---MEAHCGOTELTESLTKTAFOAHTPAQKASVLAFLINELACSKSVVSEIDKNIDYMSN 1044
Db 701 DACHELRLSNPLVKKLSSTSVYDLTPGERKMKILHALCGKLL---TLVSTDFIEDYVDI 757
Qy 1045 LRRDKWV-----EGKLR-----KURIIHAKTKGRTDTSGG 1075
Db 758 LRQAQBFRELKABQHRKEREAAAARIKRKEELKQEQKWKKEQKELKDEQORNSTAD 817
Qy 1076 IDLGEEOHPLGTPFGKRKRKKGSDSYDDDDDDDDDDQDEDEDEDEDEDEDEDEDEDEDEDE 1131
Db 818 ISIGEIE-----REDFDTSESQKDELQDMFTDEDDDPGSHKRGRRGK 864
Qy 1132 KTD-----ICEDEGDQAAASVEELEKQIEKLSKQSQOYRRKLQFPA 1172
Db 865 RGQNGFKEFTQEQINCVTRELLTADDEE---ALKQEHQKEXELKELKQS-----A 913
Qy 1173 SHLSRVMFGPDYRRRYRILPROCGIFVEGMESEGLEETAKEREKELKKAESVQIKKEEM 1232
Db 914 IACTNIFPLGRDRMYRRYWIFFSLPGLFIE--EDYSGLTEDHLLPRPSSFNQNVQSDPQ 971
Qy 1233 PET-SGDSLNCSTHDCQKEDLKEKDNINFLQKPGSFKSLKLELVAKMPPSEVWTP 1291
Db 972 VSTKTGEPLMBESTSNIDQG-----PRHVSQVLP 1000
Qy 1292 KPNAGANGCTLSYQNSGKHSLGSVQSTATQSNVEKADSNLNTFNTGSSGPGK--FYSPLPN 1349
```

Db 1001 KP-----VHK-----PNCWFYSSCEQ 1017
QY 1350 -DQLAKTLPEKNQWFSLLPRTPCDDTSLTHADMSTAS-----LVTPQSPPSK----- 1397
Db 1018 LDOLIEALNSRGR-ESALKETLLQESRI CAQLARFSEKHFSDKQPSDKPYSRGR 1076
QY 1398 -SPSTPAPLGSAQNPGLNFPALSPLOVKGVSMMGLQFCGWTGVTNIPFTLSVP 1456
Db 1077 SSNAYDPSQCAEKQLELRDLFL-----DIEDRIYQG 1110
QY 1457 SFGSGLGSEGNFSLTNSVASSSESPPQNEKATSAQPAAVEKVPDPFSPKPIPE 1516
Db 1111 TIG-----AIKVTDRHIWRS----- 1125
QY 1517 EMQFGWRIIDPEDLKALIKVLHRLGIREKALQKQIQLHLDYITQACLKNKDVAIIELE 1576
Db 1126 ALESGRYELSEN-----KENGIIKTV-----NEVVEWEIDE 1159
QY 1577 NEENQVTRDIVENWVSVEQMEMDLSVLQOVEDLERRVASASLOVKGWMCPEPASEREDL 1636
Db 1160 QTKVIVKORLLGIKTETPTSTVSTNASTPOSVSVVHYLAMALFQIE----- 1205
QY 1637 VYFEHKSFTKCKEHGDEFTGDESSAHALERK-SDNPLDIAVTRLA----- 1695
Db 1206 -----QGLERRFLKAPLD-----ASDSGRSYKTVLDR- 1232
QY 1696 IAPGLRVWRRLSEARSAAQVALCIQLOQKSIAWEKSIKMKVYCOICRKGDNELLLCDG 1755
Db 1233 -----WRESLSSASLSQVFLHLSLDRSVIWSKSIILNARCKICRKGDAENMVLCG 1285
QY 1756 CDKGCHTYCHRPKITIPDGMWFCACIAKASQTLKIKKLVKGGKTNESKK----- 1808
Db 1286 CDRGHTTYCVRPKATVPEGDFCECPKQ-----RCRLSFRQPSLESDEVEDSWG 1340
QY 1809 GKXVTLTGTE-----DED----- 1822
Db 1341 GEDDEVDGDEERGQSEEEYEVEQEDDDQSEEVSLPKRGRPVRLPVKTRGKLSGFS 1400
QY 1823 -----SASTSSSL----- 1830
Db 1401 SRGQOQEPGRYPSRQSQTPKTTVSKTGRSLURKINSAPPTTKSLRTASRSTRHSHGLP 1460
QY 1831 -----KRGNKD-----LOKRMKEENTSIN-----LSKQSF 1856
Db 1461 QADVVELLSPPRRKGRKSNANTPNSPNFPNFVIAATKSSEQSRSVNIASKLSQSE 1520
QY 1857 TSVKPKKDDSK-----DLALCSMLITEMETHEDAWPFLLPVNLKLV 1898
Db 1521 SKRRCKRQSPSPVTLGRRSSGRQGGVHLSAFEQLVVELVRHDDSWPFLKLVSKIQV 1580
QY 1899 PGYKVKIKPMDPSTIREKLSGQYPNLETALDVLVDFNCETFNEDDDSDIGRAGHNR 1958
Db 1581 PDYDIKKPIALNIIREKVNKCEYKLASEFIDDIELMFNSCFNPRNTSEAKATRLQ 1640
QY 1959 KYF 1961
Db 1641 AFF 1643

RESULT 15
US-09-270-767-45282
; Sequence 45282, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 45282

; LENGTH: 800
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
; FEATURE:
; OTHER INFORMATION: xaa means any amino acid
US-09-270-767-45282

Query Match 6.4%; Score 645.5; DB 4; Length 800;
Best Local Similarity 31.5%; Pred. No. 3.5e-37;
Matches 153; Conservative 64; Mismatches 109; Indels 159; Gaps 12;

QY 1589 NWSVEEQAMEWMDLSVLQOVEDLERRVASASLOVKGWMCPE-----EPASE----- 1632
Db 26 DMN-PKVAKRVELALIEQLESLEKVASASMLQKWLQFNVESELTLDSQSDVTEEDFV 84
QY 1633 -----REDLVVFEHKSFTKCKEHGDEFTG-----EBSAHA----- 1665
Db 85 SIIPMIRERIIDLEANIERRYLKPPLGSGQTGDAHLAVTAQNOHTTQTQNSASAAAYLLQ 144
QY 1666 -----LERKSDNPLDIAVTRLA----- 1682
Db 145 MQQQQQQQLAQQQQQQQGGSGAGNSLNPSSFNERTMALAAAAAASGPGNATGVANSVV 204
QY 1683 -----DLERNIERIEEDIAPLRVWRRLSEARSAAQVALCI 1720
Db 205 AGATPCESGSGEPNSGNASPNASNCDSRDEKVEQ-IPKGLVQMRDAVSRSHTTAQLAMAL 263
QY 1721 QOLOKSIWEKSIKMKVYCOICRKGDNELLLCDGCKGCHTYCHRPKITIPDGMWFCP 1780
Db 264 VYLESCVAVDWSIMKANCQFCTSGENEDKLLCDGCKGYHTYCFPKRMDNIPDGDWICY 323
QY 1781 ACIAKASQTLKIKKLVKGGKTNESKKGVTLTGDTEDDESASTSSSLKRGKDLQKR 1840
Db 324 ECVNKA-----TNE-----RKIVCG-----GHRPSPVG 347
QY 1841 KMEENTSINLSQESFTSVKPKRDDSXDALCSMLITEMETHEDAWPFLLPVNLKLVPG 1900
Db 348 KM---IYCDLC-----PRAYHADCLAVCKTLGEMELHEDSWPFLPVNTKQFPT 394
QY 1901 YKVKIKPMDPSTIREKLSGQYPNLETALDVLVDFNCETFNEDDDSDIGRAGHNRKY 1960
Db 395 YRKIIKTPMDLSTIKKLLQDLISYKTRDFCVDVROI FDNCFMNFEDDDSPVGKAGHMRKF 454
QY 1961 FEKKW 1965
Db 455 FESRW 459

RESULT 16
US-09-418-710-69
; Sequence 69, Application US/09418710
; Patent No. 6596482
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042001
; CURRENT APPLICATION NUMBER: US/09/418,710
; CURRENT FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 69
; LENGTH: 1525
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-418-710-69

Query Match 5.3%; Score 538.5; DB 4; Length 1525;

Best Local Similarity 18.1%; Pred. No. 3.9e-29;
Matches 353; Conservative 253; Mismatches 604; Indels 745; Gaps 64;

```
QY 189 KPLSIVNAQKETYMKLVIPSPDVLKAGNKNTSESSLLTSELRSKREOY-----KQA 241
Db 8 KPFLVNPPLGEEPP-FTIPH-----TOEAFRTREYEARLERYSERIWTCKST 55
QY 242 FPSQLKKQESSKSLKVVTAALSNPKATSSSPA-HPKQTLNHNHPFLTNALLGNHQPNG 300
Db 56 GSSQLTHKEAMEEEOEVAELK-----EFPANWEKLVLEWVHN-----TA 97
QY 301 VIQSVIOEAPIALTTKTQKSKINENIAAASSTPSSPVNLSTSGRTPGNQTPMPSAS 360
Db 98 SLEKLVDTAWLEIMTKYAVGEECDPEV-----HPL-----EKVDEE 146
QY 361 PILHSQGEKAVSNVNPVKTHSHHPAKSLVEQFRGTDSIPSKSDSENEDEEEDDE 420
Db 125 -----GKEKML--KVKTIVKI--HPL-----EKVDEE 146
QY 421 EDEBEDDDSDSDSQESDSESDTSGSEEDDDKQDESQSD-----T 467
Db 147 ATEKSDGACDPSDSDKENSQIAODHOKKTVVKEDEGRRESINDRARRSPKLPISLK 206
QY 468 EGE-----KTSMKL---NKTTSVKSPPMSLGTGHSTPRNLHIAT----- 503
Db 207 KGERKWAPPKFLPHKYDVKLQNEDKIISNV--PADSLIRTPPNKEIVRYFIRHNALRA 264
QY 504 -----APGSAAPALCSQSPPA-----FLGTSSSTL---TSSPHSGTSKRRVVT 544
Db 265 GTGENAPVWVDELVKYSLSKFSDFLLDPYKMTLNPSTKRKNTGSPDRKPKSKK-T 323
QY 545 DERELRIPLEYQWQRETRIIRNFGRLQGEVAYAPCGKKLRQYPEVIKVLNRNGIMDISR 604
Db 324 DNSSLSSEL-----NPKLMCHVHLKKSLS----- 347
QY 605 DNFSFSAKIRYDGFYEARDGPOEMOWCL-----LKEEDVIPRIMAGRRGPPNPDR 657
Db 348 -----GSPKVKNSKNSPEEHLKEMKMWSPNKLHTNPHPK-----KGP-PAK 392
QY 658 QRAREESRMRKGRPPNVGNAEFLDNADAKLLRKLQAEIARQAAQILKRLKQKQQA 717
Db 393 KPGKSHDKPLKAKGRSGKI-----LNQKST 418
QY 718 RVAKKAKQQAIAAAEERKKEQIKIMKQEKIKRIOQ-----IRMEKEL--RAQ 766
Db 419 GNSKSPKX--GLKTPKTQKQMTLDMAKGTQKMTAPRNSGGTPRTSSKPHKLLPPAAL 476
QY 767 QILEAKKKKKEE-----AANAKLLEAEKRIK-EKEMR---ROQAVLLKHOER-- 809
Db 477 HLIAYYKENDKREDKRSALSCKVISKTARLLSSEDEARLPEELRSILVQKRYELLEHKKWA 536
QY 810 ---ERRRHHMLKAMAEARKKAEKERLUKQKREDEKRLNKERKUEQORLELEMAKELKKP 866
Db 537 SMSERQKRYLKKREELKKLKEKAKERREKEMLEKQKRYEDQEL---TGKNL--- 590
QY 867 NEDMCLADQKLPPELPRIPGLVSGTSDCLMVVQFLNFKVLGFPDVIIDVPLVLQ 926
Db 591 -----PAFRLVDTPF-GLPNTLFGDVAMVVEFLSCYSGLLLPDAQYPIAVS-IM 638
QY 927 EGLNIGDSMGVEQDILLVRLLSAAVCDPGLITGYKAKTALGHEHLNVGNVRDNVSEILQI 986
Db 639 EALADKGGFL-YLNRVILLIQTLLQDE-IAEDY---GELGNKLSEIPLTHLSVSELVRL 693
QY 987 FM-----EAHGGOTELTESLTKTAFOAHTPAQKASVLAFILN 1023
Db 694 CLRRSDVQESBGSDTDNDKSDAAFEDEVQDEFLEKLETSEFFELTSEEKQLQILTALCH 753
QY 1024 ELACSKSVSEIDKXIDYMSNLRRDKWV-----EGKLRLRIIHAKKTGKDTSGGIDLG 1079
Db 754 RILMTYSVDHMETRQQAELKWLKRLAVLKEENDKKAERKQKREMAKNKENGKVBNG 813
QY 1080 EQEHPLGTPTPGRKRRRKGGSDDYDDDDDDSD-----QGEDEDEEDKE 1126
```

```
Db 814 -----LG-----KTDRKKRIVKPEPQVDTEAEDMISAVKSRRLLAIOAKKEREIQE--R 860
QY 1127 DQKGTKDIDCEDEGDOAASVELEKQIEKLSKQSQYRKLKFDASHLSLSVMFGPPRY 1186
Db 861 EMKVKLERQAEERIRKHAAAEK-----AFQEGIAKAKLVMRPTIGTDNRN 907
QY 1187 RRRYUIL-PRGGGIFVEGMESGEGLEBIAKEREXLKAESYOIKEEMFETSG---DSLNC 1242
Db 908 HNRVWLSDEVPLGFIE-----KGVVHDSIDY 934
QY 1243 SNTDHCOKEDLKEDKDNINFLQKPGSFSKLSKLELVAKMPPSEVMTPKNAGANGCTL 1302
Db 935 RFNHC-----KDHT-----VSGCED 950
QY 1303 SYQNSGKSLGSVOSTATQSNVERKADSNNLNFTSGSGPKFYSPLPNDQLLTKLTKENRQ 1362
Db 951 YCPRSKANLGNASMTQHG-----TATE----- 975
QY 1363 WFSLLRPTPCDDTSLTHADMSLASLVTPSQSPSPSPPTAPLGSSAQNPVGLNPPFALS 1422
Db 976 ----- 975
QY 1423 PLQVKGVMGMLQFCGWPTGVTSNIPFTLSVPSLGLSGLSEGNNSFLTSNVASSKS 1482
Db 976 ----- 975
QY 1483 ESPVPQNEKATSAQPAAVEVAKPVPFSPKPIPEEMQFGWRIIDPEDKALLKVLHLRG 1542
Db 976 -----VAVETT-----TPQGNLWFLCDSQKDELLNCLHPQG 1010
QY 1543 IREKALOKIOK-HLDVITQACLKXK-DVAIIELENBEE--NOVTRDIVENWSVEEQAME 1598
Db 1011 IRESQLEKLEKRYQDIHSHLARKNLGLKCDGNQELNFLRSDLIE---VATRLQK 1067
QY 1599 MDLSVLOQVEDLERVAS-----ASLQ-----VKGWMCPEPASEREDLVIFE 1640
Db 1068 GGLGVBEETSEFEARVISLEKLDKDFGECVIALQASVIKFLQGFMAPKQ----- 1116
QY 1641 HKSFTKLCKEHGFTGEDESSAHALERKSDNPLDIAVTRLADLERNIERRIEEDIAFGL 1700
Db 1117 -----KRRKLOSDSAKTEEVDEEKGVE--EAKVASAL 1148
QY 1701 RWRRALSEARSAQAVALCICLOQKSIKAWKSIKMYVCOICRKGDNELLLLCDGCKGC 1760
Db 1149 EKWTAREATQFSRHHVLLGLDACLKWDNSAENARCKVCPKKGEDDKLLILCDECNKAF 1208
QY 1761 HTYCHRPKITTIPOGDFPCACIAKASQTLKIKLHVKKKTNSKKGKVTLTGDTED 1820
Db 1209 HLFCLRPALYVDPDGEWQCPACQAPATA-----RRNSRGRNYTEESASEDESEDEE 1261
QY 1821 EDS-----ASTSSSLKRGKNDLQKRWKWEENT-SIN 1849
Db 1262 EEEEEEEEDYEVAGLRLRPRKTRIRGHKSVIPPAARSGRRPGKKPHSTRSQKAPPVD 1321
QY 1850 LSKQESFTSVKKPRDDSKDIALCSMLITMETHEDEAMPFLLPVNLKLVPGYKVKIKKPM 1909
Db 1322 AEVDLVLQTKRSRRQSLQKCEELHKVYKRFSPFPFVTRDEADYDYVITHPM 1381
QY 1910 DFSTIREKLSGGQPNLETFPALDVLRLVFNDCETFN 1944
Db 1382 DFQIVQNKSCSGSYRSVQEFLLTDMKQVFTNAEYVN 1416
```

RESULT 17
US-09-839-479-68
; Sequence 68, Application US/09839479
; Patent No. 6727222
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042002
; CURRENT APPLICATION NUMBER: US/09/839,479
; CURRENT FILING DATE: 2001-04-20

Db 1322 AEVDVLQTKRSSRQSLQKCBELHKIVKVRFSWPFREPVRDDEADYDVTIHPM 1381
Qy 1910 DFTIREKLSGOYENLETFEALDVLVDFDNCETFN 1944
Db 1382 DFTQVONKCSGYSRVSQEFELTDMQVFTNAEVYN 1416

RESULT 18
US-09-949-016-7037
; Sequence 7037, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7037
; LENGTH: 1540
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-7037

Query Match 5.3%; Score 536; DB 4; Length 1540;
Best Local Similarity 18.0%; Pred. No. 6e-29;
Matches 348; Conservative 254; Mismatches 584; Indels 744; Gaps 66;

Qy 221 SEESLLTSELRSKREQV-----KQAPPQOLKKQESSKSLKKVIAALSNPKATSSSPA 273
Db 85 TQEAFTREEYEARLERISRIWTCKSTGSQLTHKEAWEEQEVAELK-----EEFPA 139

Qy 274 -HPKOTLENNHPNPFLLTNALLGNHOPNGVISOVIOEAPLALTTTKMOSKINENIAAASS 332
Db 140 WYELKLVLEWVHN-----TASLEKLVDTAMLEIMTKYAVGECDFEV----- 181

Qy 333 TPFSSPVNLSTSGRTPGNTQTPMPSASPIILHSQCKEKAVERNPNVKTQHHSHPAKSLV 392
Db 182 -----GKERKL--KVKIVKI-----HPL----- 197

Qy 393 EQFRGTDSIPSSKSDSDNEDEDEDEDEDEDEDESDSDSDSDSDSDSDSDSDSDSDSD 452
Db 198 -----EKVDEATEKSDGACDSPSSDKENSQIAQDHQKKE 235

Qy 453 EDDDDKQDESDD-----TEGE-----KTSMKL---NKTTSVKS 485
Db 236 VVKEDEGRRESINDRARRSPKLPSTSLKKGERKWAAPPKFLPHKYDVKLQNEDKIISNV-- 293

Qy 486 PMSLTGHTSPNHLHIK-----APGSAPALCSESQSPA----- 520
Db 294 PADSLIRTPPNKEIIVFYIRHNALRAGTGENAPWVVEDELVKKYSLPKSFDFLLDPY 353

Qy 521 -FLGTSSSTL---TSSPHSGTSKRRVVTDERELRIPLEVG-WORETRIRNFGG---RLQG 572
Db 354 KYMTLNPSTKTKNTGSPDRKSKSK-TDNSLSLSPFLNPKLVCHVHLKKSLSGSPKVKYN 412

Qy 573 EVAYYAPCGKKLRQYPIVYKILSRNGIMDI GRDNFSPSAKIRVGDFYBARDGPQBMQWCL 632
Db 413 SKNSKSP---EEHLEEMKWMSPN-----KLHT-NFHIPKKGK----- 446

Qy 633 LKEEDVIPRIAMEGRGRPPNPQORARESRMRRRKRPNVGNAEFLONADAKLLRK 692
Db 447 -----PAKKPGKSDKPLKAKGSKGI----- 468

Qy 693 LQAQEIARQAQIKLLRLKQEQARVAKKQQAQIWAABEKRKQEQIKIMKQEQEKIK 752

Db 469 -----LNGQKSTGNSKSPK--GLKTPKTKMQMTLLDMAKGTQKWT 508
Qy 753 RIQQ-----IRMEKEL--RAQOILEAKKKKKE-----AANAKLLEAEKR 790
Db 509 RAPRNSGGTPTSTSKPHKHLPPAALHLJAYYKENKDRDKSALSVCVISTARTLSSSDR 568

Qy 791 IK-EKEMR---ROQAVLLKHOER-----ERRROHMLMKAMEARKKAEEKERLQOEKRDE 841
Db 569 ARLPBELRLSVOKRYELLEHKRWASMSSEBQKEYLKKREELKKLKEKAKERREKEML 628
842 KRLNKKERKLEQRLELEMAKELKKPNEDMCLADQKLPFELPRIPCLVLSGTFSDCLMVV 901
Db 629 ERLEKQKRYEQEL---TGKNL-----PAFRLVDTPE-GLPNTLFGDVAMVV 671

Qy 902 QFLNFKVLGFDVYNDVNPVLSVLOEGLNLIGDSMGEVQDILLVRLLSNAVCDPGLITQYK 961
Db 672 EFLSCYGLLLPDAQYPTITAVSLMEALSADKGGFL-YLNRVLVILLQTLQDE-IAEDY- 728

Qy 962 AKTALGHEHLLNVGRDNRVSEILOIFM-----EAHCGQTELT 998
Db 729 --GELGMKLSIPLTHSVSELVRLCLRRSDVQSESGSDTDNKKDSAAFEDNEVQDFL 786

Qy 999 ESKTKAFOAHTPAQKASVLAFLINELACSVSVSEIDKNIDYMSNLRDRKVVV---EG 1054
Db 787 EKLETSFFELTSEKQLILTALCHRILMTYSVQDHMETRQOMSAAELWKERLAVLKEBD 846

Qy 1055 KLRKLRIHAKTKGRTSGDILGEBQHPLGTPGKRKRKRGDSDYDDDDDDSD- 1113
Db 847 KKRAEKQKRKEMAKENGKVENG-----LG-----KTDKKEIKVFEPQVDTAEDM 895

Qy 1114 -----QGEDEDEDEKEDOKGKKTIDICEDEDEGQAAASVEELEKQTEKLSQ 1161
Db 896 ISAVKSRRLLAIAKKEREIQE--REMKVKLERQAEERIRKHKAABK----- 942

Qy 1162 QSOYRRKLFDAHSLRSVMFGPDYRRRYWTL-PRCGGIFVEGMESGEGLEIEIAKEREKL 1220
Db 943 --AQEGIAKAKLVMRRTPIGTDNRHNRVYLFSDVPGFLFIE----- 982

Qy 1221 KKAASVOIKEEMFTSG---DSLNCSTYDHCQEKEDLKEKONTNLFLOKPGSFKLSKLL 1277
Db 983 -----KGWVHDSIDYRFNHHK-----KDHT---VSGDEDYCPRSK-- 1014

Qy 1278 EVAKMPPESEVMTPKPNAGANGCTLSYQNSKHSLSGVSQSTATQSNVKEKADSNLNTGS 1337
Db 1015 -----KANLGKVA-----SMTQHGATATEVAE----- 1037

Qy 1338 SGPGKFSPLPNDQLKTLTEKNRQ--WFSLLPRTPCDDTSLTHADMSTASLAVTPQSOPP 1395
Db 1038 -----TTTPKQGNLWFL-----CD----- 1052

Qy 1396 SKSPSPAPILGSSAQNPVGLNPFALSPLOVKGVSGMMGLQFCGWPTGVVTSNIPFTLSV 1455
Db 1053 ----- 1052

Qy 1456 PSLGSLGLSGNGSNFLTSNVASSKSESPPQNEKATSAQAPAAVEVAKPVPDFSPKPIP 1515
Db 1053 ----- 1052

Qy 1516 BEMQFGWWRITDPEDLKALLKVLHLRGIREKALQKIQK-HLDYITQACLKNK-DVAIIE 1573
Db 1053 -----SQKELDELLCNLHPQIGRESQLEKRYQDIIHSHLARKPNLGLKS 1101

Qy 1574 LNEEBE-NQVTRDIVENWSVEEQAMEMDLSVLOQVEDLERRVAS-----ASLQ 1620
Db 1102 CDGNQELNLFNRLDIE---VATRLQKGLGYVEETSEFEARVISLEKLDKDFGECVIALQ 1158

Qy 1621 -----VKGMWCEPASEREDLVYFBHKSFTKLCKHDGDEFTGEDESSAHALKERKSDNP 1673
Db 1159 ASVTKKFLGGMAFKQ-----KRRKLOS 1181

Qy 1674 LDIAVTRLADLERNIERRIEEDIAPLGRVWRRLSEARSAAQVALCIOOLQKSTAWEKSI 1733

Db 1182 EDSAKTEEVDEBKWVE--EAKVASALEKWKTAIRBAQTFSRMHVLLGMLDCAIKWMSA 1239
Qy 1734 MKVYCOICRKGONBELLCDCKGCHTYCHRPKITITPIDGDFCPACIAKASGOTLKI 1793
Db 1240 ENARCKVKCKGEGDKILDCBNKAFHLFCLRPALYEVDPGEWQCPACQAPATA----- 1293
Qy 1794 KKLHVKGKKTNSKKGKVTLTGDTDEDS :|||: 1823
Db 1294 -RRNRRGRNYTESASE-----DSEDDDEDEEBEEDYEVAGLRRLRPKRTIR 1346
Qy 1824 -----ASTSSSLRGKNDLQKWEENTS--INLSQESPTS VKPKRDDS KDLALCS 1874
Db 1347 GKHSVIPPAAARGRRPGKPHSTRSQPKAPPVDDAEVDVLQYKRSRROSLELQKE 1406
Qy 1875 MILTEMETHEDAWPFLPNLKLVPYKVKVKKPMDFSTIREKLSGGYQPNLETFALDVR 1934
Db 1407 EILHKIVKYRFSWPPREPVRDEADYDVITHPMDFTVQNKCSGVSRSVQBELTDMK 1466
Qy 1935 LVPDNCETPN 1944
Db 1467 QVFTNAEVYN 1476

RESULT 19

US-09-418-710-29
; Sequence 29, Application US/09418710
; Patent No. 6596482
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042001
; CURRENT APPLICATION NUMBER: US/09/418,710
; CURRENT FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 29
; LENGTH: 1531
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-418-710-29

Query Match 5.3%; Score 533; DB 4; Length 1531;
Best Local Similarity 17.9%; Pred. No. 9.8e-29;
Matches 352; Conservative 256; Mismatches 600; Indels 756; Gaps 67;

Qy 189 KPLSLVNOAKETYMKLI VPSDVLKAGNKNTSESSLLTSLRSKROY-----KQA 241
Db 8 KPFLVNLPGEEF-FTIPH-----TQFAFRTREYEARELRYSEIRIWTCKST 55
Qy 242 FPSQLKQESSKLLKVAALSNKATSSSPA-HPKQTLNHNHPFLTNALLGNHPNG 300
Db 56 GSSQLTHKEAWEEQEVALLK-----SEFPAYEKLVLVWVHN-----TA 97
Qy 301 VTQSVIOEAPLALTTTKMQSKINENIAAASPTPSSPNVLTSGRRFTGNQTPWPSAS 360
Db 98 SLEKLVDTAWEITMYAVGECDFEV----- 124
Qy 361 PILHSQGEKAVNNVNPVKTQHSHHPAKSLVEQFRGTDSIPSSKDSNDEDEE 420
Db 125 -----GKEKML--KVKIKI-----HPL-----EKVDEE 146
Qy 421 BEDEDDDEDDSDQSESDNSDSTEGSEEDDDDKQDESDSD-----T 467
Db 147 ATEKSDGACDPSDDKENSQIAQHQKKEITVKEDEGRRESINDRARRSPKLP TSLK 206
Qy 468 EGE-----KTSMKL--NKTTSVKSPMSLTGHSTPRNLHIK----- 503

Db 207 KGERKWAAPPKFLPHKYDVKLQNEDKIISNV--PADSLIRTERPPNKEIVRYFIRNALRA 264
Qy 504 -----APGSAPAALCESQSPA-----FLGTSSSTL-----TSSPHSGTSKRRVVT 544
Db 265 GTGENAPWVVEDELVKYSLSPSKFSDFLDPYKMTLNPSTKRKNTGSPDRKPSKSK-T 323
Qy 545 DERELRIPLEYG-WQRETRIRNFG-----RLOGEVAYYAPCGKKLRYQYDEVIKYLSRNGIM 600
Db 324 DNSLSLSPNPKLWCHVHLKESLGSPLKVNKNKNSKSP-----BEHLEEMKMMSPN--- 376
Qy 601 DISRDNFSFAKIRVGDFYEARDGPQEQMOWCLLKEEDVIPRIAMEGRGRPPNPDRORA 660
Db 377 -----KLHT-NFHIPKGP-----PAKPG 395
Qy 661 REESMRRRKGRPNVGNABFLDNADAKLLRKLQAEIARQAQIKLLRKLQKQBARVA 720
Db 396 KHSKPLKAKGRSGI-----LNGQKSTGNS 421
Qy 721 KEAKKQQAIIAAAEKRRKQEQIKIMKQOEKIKR100-----IRMEKEL--RAQOIL 769
Db 422 KSPKK--GLKTPKTKMQMTLLDMAKGTQKTRAPRNSGGTPRTSSKPKHKLPPAALHLI 479
Qy 770 EAKKKKEE-----AANAKLLAEAKRIK-EKEMR---RQQAULLKHOER----- 809
Db 480 AYYKENKDREDKRSALSCVISKTARLLJSSDRARLPEELSLVQKRYELLEHKRWASMS 539
Qy 810 ERRRQHMLMKAMEARKKAEKERLKBKRODEKRLNKERKLEQRRLLEMAKELKKPNED 869
Db 540 EEOKEYLKKREELKKLKEKAKEREKEMLEKQKRYEDQEL---TGKUL----- 590
Qy 870 MCLADQKPLPELPRI PGLVLSGSTFSDCLMVVQVLPNFGKVLGFDVNDVNLVQLBGL 929
Db 591 -----PAFLVDTPE-GLPNTLFGDVAMVVEFLSCYSGLLLPDAQYPIAVSLMEALS 642
Qy 930 LNIQD--SMGEVQDILLVRLLSAAVCDPCLITGYKAKTALGEHLNVGNRDNVSIILIF 987
Db 643 ADKGFLYLNRLVILLQTLLOLQDEIADY---GELGMKLSIPIULHSVSELKLC 699
Qy 988 M-----EAHCGQTELSKTKAFQAHTPAQKASVLAFLINE 1024
Db 700 LRRSDVQEESESGSDTDNDKDSAAFEDNEVDQEFLEKLETSEFFELTSEKQLITALCHR 759
Qy 1025 LACSKSVVSEIDKNIDYMSNLRDRKWV---EGKLRLRIIHAKTOKRTSGDIDGE 1080
Db 760 ILMTYSVDHMETRQOMSALWELKERVLALEKENDKRAEKQKREMAKNGKENGKENG- 818
Qy 1081 EQHPLGTPTGKRRRKGSDSDYDDDDDDSD-----QGEDEDEDEEDKED 1127
Db 819 -----LG-----KTRKKRIVKFPQVDTEADMISAVKSRRLLAIAQAKEREIQE--RE 866
Qy 1128 QKGKKTIDICEDEGDOAASVEELEKQIEKLSKQSOVRRKLFDAHSLSRSMFGPDRYR 1187
Db 867 MKVLEKQAEERTRKHAAAEK-----AFQEGIAKAKLVMRRTPIGTDRNH 913
Qy 1188 RRYWIL-PRCGGI FVEGMESGEGELEIAKEREKLLKAEVSQIKEEMFTSG---DSLNC 1243
Db 914 NRYWLFSEVPGFLIE-----KGWVHDSIDYR 940
Qy 1244 NTDICEQEKEDLKEKONTNLFLOKPGSFKLSKLLKLVAKMPPESEVMTPKPNAGANGCTLS 1303
Db 941 FNHHC-----KDHT---VSGDEDYCPRSK-----KANLGKNA----- 969
Qy 1304 YQNSGKHSLSGVSQSTATQSNVEKADSNLFPNTGSSGPGKFYSPLPNDQLLTLTKNRQ- 1362
Db 970 -----SMNTOHGTATEVAE-----TTTPKQCOQN 993
Qy 1363 -WFSLLPRTPCDDTSLTHADMSTASLVTPQSPFSKSPPTAPPLGSAQNPVGLNPPAL 1421
Db 994 LWEL-----CD----- 999
Qy 1422 SPLQVKGVSMMGLQFCGWPTGVVTSNIPFTLSVPSLSGSLGSEGNNSFLTNSVASSK 1481
Db 1000 ----- 999

Db 914 NRYWLFSDVPGLFIE-----KGWVHDSIDYR 940
QY 1244 NTDHCEQEDLKEKONTNLFLOKPGSFSLKSLLEVAKMPPEVMTPKPNAGANGCTLS 1303
Db 941 FNHHC-----KDHT---VSGDEDYCPRSK-----KANLGKNA--- 969
QY 1304 YQNSGKHSLSGVOSTATQSNVEKADSNLFWTSGSGPGKFYPLPNDQLLTLTKENRQ- 1362
Db 970 -----SMNTOHGATEVAVE-----TTTPKQGN 993
QY 1363 -WFSLLPRTPCDDTSLTHADMSTASLVTQSPPSKSPPTAPLGSQAQNPVGLNPPAL 1421
Db 994 LNFL-----CD----- 999
QY 1422 SPLQVKGVMGLOFCQGPTGVVTSNIPFTLSVPSLGSGLSEGNGNSFLTNSVASSK 1481
Db 1000 ----- 999
QY 1482 SESVPQNEKATSAQPAAVEVAKPVDFFSPKPIPEMQPGWRIIDPEDLKALLKVLHLR 1541
Db 1000 -----SKELDELLNCLHPQ 1014
QY 1542 GIREKALOKIOK-HLDYITOACLKXK-DVAIIELNENEE--NOVTRDIVENWSVEEQAM 1597
Db 1015 GIRESQLKERLEKRQODIIHSHLARKPNGLKSCDGNQLNLFRLSDIE---VATRLQ 1071
QY 1598 EMDLSVLOQVEDLERRVAS-----ASLQ-----VKGMWCPPEASEREDLVYF 1639
Db 1072 KGLGVVBETSEFEARVISLEKLKDFGECVIALQASVIKPIQGFMAPKQ----- 1121
QY 1640 EHKSTFKLCKEHDGFTGEDESSAUALERKSNPDIATVRLADIERNIERIEBEDIAPG 1699
Db 1122 -----KRLKQSEDSAKTEVEDEKQWVE--EAKVASA 1152
QY 1700 LRVWRRALSEARSAQVALCLOQLOKSTAWKSIKVVYCOICRKGDNELLLCDGCKG 1759
Db 1153 LEKWTAIREAQTFRMRHVLGNLADACIKWDSANARCKVCPKGEODDKLILCDECNKA 1212
QY 1760 CHTYCHRPKITIPDGMFCPACIAKASQOTLKI KKLHVKGKKTNESKKGKVTILTGDT 1819
Db 1213 FHLFCLRALYVDPGEWQCPACQATA-----RNSRGRNYTEESASE-----DSE 1259
QY 1820 DEDS-----ASTSSSLKRGNKLOKRM 1842
Db 1260 DDESDEEBEEDYEVAGLRLPRKTIRGKHSVIPAARSGRPGKPHSTRRS 1319
QY 1843 BENTS--INLSQESFTSVKPKRDDSKDLALCSMILTEMETHEDAWBFLLPVNLKLVPG 1900
Db 1320 QPKAPPVDDAEVDELVLQTRSSRSQSLQKCEELHKKIVKYSWPFREPVTREABD 1379
QY 1901 YKKVIKPMDFSTIREKLSSGQPNLETFALDVLVLFONCETFN 1944
Db 1380 YYDVITHPMDFTQVQNKCSGYSRVSQBEFLDMKQVFTNAEYVN 1423

RESULT 21

US-09-418-710-27
; Sequence 27, Application US/09418710
; Patent No. 6596482
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042001
; CURRENT APPLICATION NUMBER: US/09/418,710
; CURRENT FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JF98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 27
; LENGTH: 1527
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-418-710-27

Query Match 5.2%; Score 532; DB 4; Length 1527;
Best Local Similarity 18.0%; Pred. No. 1.1e-28;
Matches 354; Conservative 255; Mismatches 597; Indels 756; Gaps 68;

QY 189 KPLSLVNAQKKTMYKLVIPSPDVLKAGNKNTSESSLLTSELRSKREOY-----KQA 241
Db 8 KPFVLNPLGEBEP-FTIPH-----TQEAFTREYEARELRYSERIWTCKST 55
QY 242 FPSQLKKQESSKSLKVVIAALSNPKATSSSPA-HPKQTLNHNHPNPFLLNALLGNHPNG 300
Db 56 GSSQLTHKEAWEERQEAELLK-----EFPANWEKLVLEMVHN-----TA 97
QY 301 VIQSVIQEAPLALTTKTKQSKINENIAAASSTPFSSPVNLSSTGRRTPGNQTPVMPAS 360
Db 98 SLEKLVDTAWLEIMTKYAVGEECDFEV----- 124
QY 361 PILHSQKKEKAVSNVNPVKTQHHSHPAKSLVEQFRGTDSDIPSSKQSEDSNEDEEEDDE 420
Db 125 -----GKERML--KVKIVKI---HPL-----EKVDEE 146
QY 421 REDEEDDEDDSDSQSESDSNESDTEGSEEDDDDKDQDESDS-----T 467
Db 147 ATEKSDGACDSPSSDKENSQIAQDHQKQKTVKVEDGRRRESINDRARRSPRLKPTSLK 206
QY 468 EGE-----KTSMKL---NKTTSVKSPKSLTGHSTPRNLHTAK----- 503
Db 207 KGERKWAPPKPLPHKYDVKLQNEDKIISNV--PADSLIRTERPPNKEIVRYFIHNALRA 264
QY 504 -----APGSAAPALCSESQSPA-----FLGTSSSTL---TSSPHSSTSKRRVVT 544
Db 265 GTGENAPVWVEDELVKKISLPSKSFDFLLDPYKMTLNPSTKRWKNGSPDRKPSKSK-T 323
QY 545 DERELRIPLEYG-WQRETRIRNFGG---RLQGEVAVYAPCGKCLQRQYPEVIKLYSRNGIM 600
Db 324 DNSSLSPFLNPKLCHVHLKKSLSGSLPKVKNKSKSP---BEHLEEMKQWSPN--- 376
QY 601 DISRDNFSFSAKIRVGDYFYEARDGPOEQMWCLKEEDVIPRIAMEGRGRPPNDRORA 660
Db 377 -----KLHT-NFHIPKGGP-----PAKPG 395
QY 661 REESMRRRKGRPPNVGNNAEFLDNADAKLLRKLQAEIARQAQIKLLRKLQKQSOARVA 720
Db 396 KHSKPLKAKGRSKGI-----LNGQKSTGNS 421
QY 721 KEAKKQQAIAAAEKKRQKEQIKIMKQOEKIRIQO-----IRMEKEL--RAQOIL 769
Db 422 KSPKK--GLKTPKTKMQMTLLDMAKGTQKMTAPRNSGGTPTRTSSKPKHLPPAALHLI 479
QY 770 EAKKKKKEE-----AANAKLLEAEKRIK-EKEMR---RQOAVLLKHQER----- 809
Db 480 AYYKENKDRKRSALSCVISKTARLLSSEDRARLPEELRSLVQKRYELLEHKKRWASMS 539
QY 810 ERRQHMLMKAMEARKKAEKERLKOKEKDEKLEKLEQRLELEMAKELKKNED 869
Db 540 EQQRKEYLKKREELKKLKEKAKERREKEMLEKOKRYEODEL---TGNNL----- 590
QY 870 MCLADQKPLPRLPIPLGLVLSGSTFSDCLMVVQVFLRNFVKVLGFDVNDVNLVLOEGL 929
Db 591 -----PAFLVDTPE-GLPNTLFGDVAMVVEFLSCYSGLLPPDAQYPIITAVSLMEALS 642
QY 930 LNIQDSMGEVODLVRLLSAAVCDPGLITGYKAKTALGEHLNNGVNRDNVSEILQIFM- 988
Db 643 ADKGGFL-YLNRVLVILLQTLQDE-IAEDY---GELGMKLSIPLTLHVSSELVRLCLR 697
QY 989 -----EAHCQGTLETSLTKTAKFOAHTPAQKASVLAFINELA 1026
Db 698 RSDVQEESESGDSTDNDKDSAAAFEDNEVQDEFLEKLETSEFFELTSEEKLIQILTALCHRL 757

Db 480 AYYKENKREDKRSALSCVISTARLLSSDRARLPBELRSVLQKRYELLEHHKKWASMS 539
Qy 810 ERRQRHMLMKAMEARKKAEERKRLQEKRRDKRLNKKERLQORLELEMAKELKPNED 869
Db 540 EQRKEYLKKREELKKERAKERREKEMLERLEKQKRYEDQEL---TGNL----- 590
Qy 870 MCLAQOKPLPELPRIPGLVLSGTSFSDCLMVVQFLRNFGKVLGFDVNDVNPVLQOGL 929
Db 591 -----PAFRLVDTPE-GLPNTLFGDAMVVEFLSCYSGLLPDAQYPIITAVSMEALS 642
Qy 930 LNWIGSMGVQDILLVRLSAACVDPGLITGYKAKTALGHLNAGVNRDNDVSEILQIFM- 988
Db 643 ADKGGL-YNRVVLLOTLQDRE-IAEDY---GELGWLKSEILTHSVSELVRLCLR 697
Qy 989 -----FAHCGQTELSLTKAFAOHTPAQKASVLAFLINELA 1026
Db 698 RSDVQBEESGSDTDNDKSAAFEDNEVDQEFLEKLETSEFFELTSEELQILTALCHRL 757
Qy 1027 CSKSVVSEIDKXNDYMSNLRDKWV---EQLKRLRIIHAKTGKRDTSGGIDGEEQ 1082
Db 758 MTYSVDHMETRQOMSAELWKLAVLKEENDKRAEKQKREMEAKNKENGKGVENG--- 814
Qy 1083 HPLGTPGKRRRRGGSDYDDDDDDSD---QGEDDEDEEDKEDQK 1129
Db 815 ---LG-----KTRKRIIVKFPQVDTEADMISAVKSRLLAIQAKKERIQE---REMK 864
Qy 1130 GKTDICEDEGDOAASVEELEKQIEKLSKQSOVRRKLFDAHSLSRVSMGPDYRARR 1189
Db 865 VKLERQAEERIKHKAEEK-----AFQEGIAKALVMRTPIGTDRNHR 911
Qy 1190 YWIL-PROGIFVWGESGEGLEETAKERBLKKAESVOIKEEMFETG---DSLNCST 1245
Db 912 YWLFDEVPGLFIE-----KGVVHDSIDYRFN 938
Qy 1246 DCEQEKEDLKEKNTNLFLOKPGSKLSKLEVAKMPPESEWTPKKNAGANGTSLVQ 1305
Db 939 HHC-----KDHT---VSGDEDYCPRSK-----KANLGNA----- 965
Qy 1306 NSGKSLGASVQSTATQSNVEKADSNLNTGSSGPKFYSLPDLNQLLTKTEKNRQ--W 1363
Db 966 -----SMNTQHGTAETAVE-----TTTPKQOGLN 991
Qy 1364 FSLPRTPCDRLSLTHADNSTASLVTPQSPKSPSPPTAPLGSQAQNPVGLNFPALSP 1423
Db 992 FL-----CD----- 995
Qy 1424 LQVKGVSMMGLQFCGWPTGVVTSNIPFTLSVPSLGLSLGGLSEGNGSFLTSNVASSKE 1483
Db 996 ----- 995
Qy 1484 SPVPQNERKATSAQAAPAAVEAKPVPFPSPKPIPEMQFGWRIIDPEDLKALLKVLHRLGI 1543
Db 996 -----SOKELDELNCLHPQGI 1012
Qy 1544 REKALQKQIK-HLDYITQACLKNK-DVAIIELNENEE--NQVTRDIVENMSVBEQAMEM 1599
Db 1013 RESQLKERLEKRYQDIHSIHILAKPNLGLKSCDGNQELLNFRSLDLIE---VATRLQKG 1069
Qy 1600 DLSVLOQVEDLERRVAS-----ASLQ-----VKGMCPPEPASEREDLVYFEH 1641
Db 1070 GLGYVEETSEFARVTSLEKLDGFCVIALQASVIKKFLQGFMAPQK----- 1117
Qy 1642 KSFTKLCKEHDEGTEDESSAHALERKSDNPLDIAVTRLADLERNIERRIEDIAPIGLR 1701
Db 1118 -----KRRKQEDSAKTEVDEEKKWVE--BAKVASALE 1150
Qy 1702 VWRRLSEARSAAQVLCIQQLQKSIAMEKSMKYVCQICRKGDNBEELLGLDGGDKGCH 1761
Db 1151 KWKTAIREAQTSRHHVLLGLMDACIKWDMSEAENARCKVCPKKGDDKLILDECNKAFH 1210
Qy 1762 TYCHRPKTIIPDGOWFCPACITAKASGQTLTKLHVKGKKTNESKKGKVTLTGDTDE 1821
Db 1211 LFCRLPALYVEVDGEGWQCPACOPATA-----RRNSRGRNTEESASE-----DSEDD 1257

Qy 1822 DS-----ASTSSSLKGNKDLQKRKMBE 1844
Db 1258 ESDEEEEEEEBEEDEYEVAGRLRPRKRTIRGKHSVIPPAAARSRRRPGKPKPHSTRSQP 1317
Qy 1845 NTS---INLSKQESFTSVKPKRDDSKDALCSMLITEMETHEDAWPFLPYNLKLVPGVK 1902
Db 1318 KAPPVDDAEVDLVLQTKRSSRROSLELOKCEELHKIKYRFSWPFREPVTREADEYY 1377
Qy 1903 KVIKPMDFSTIREKLSGQYPNLETFALDVRVLVFNDCETFN 1944
Db 1378 DVITHPMDFOIVQNKSCGSRYSVQEFLLTDMKQVETNAEVYN 1419

RESULT 23
US-09-418-710-65
; Sequence 65, Application US/09418710
; Patent No. 6596482
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042001
; CURRENT APPLICATION NUMBER: US/09/418,710
; CURRENT FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: Fast-Seq for Windows Version 4.0
; SEQ ID NO 65
; LENGTH: 175
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-418-710-65

Query Match 4.4%; Score 445.5; DB 4; Length 175;
Best Local Similarity 53.4%; Pred. No. 6.3e-24;
Matches 94; Conservative 31; Mismatches 50; Indels 1; Gaps 1;

Qy 850 LEQRRLLEMAKELKKNEDMCLADQKPLPELPRIPGLVLSGTSFSDCLMVVQFLRNFGK 909
Db 1 LEERQKQMLEENKPKTEDMCLTDHQPDPFSRVPGLTLPSPAEDCLTIVEFLHSEFG 60
Qy 910 VLGDVNDVNLVQEGILNIGDSMCEVODLLVRLLSAAVCDPGLITGYKAKTALGEH 969
Db 61 VLGFDPKDVPSLGLVQLQEGLLCQGDLSLGEVQDILLVRLKAAHLDHDPFYSYQCSLKILGEK 120
Qy 970 LLNVGNRDNVSEILQIFMEAHCGQTELSLTKAFAOHTPAQKASVLAFLINEL 1025
Db 121 VSEPLTRDNDVSEILRCPLMAYGVEPALCDRLRTQPPQAO-PPQKAAVLAPVPHL 175

RESULT 24
US-09-839-479-64
; Sequence 64, Application US/09839479
; Patent No. 6727222
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042002
; CURRENT APPLICATION NUMBER: US/09/839,479
; CURRENT FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18

Db	41	SLNFSEGDGRRRRVLLRGRESPAAGPRYSSEGLSPSKRRRLSRWNHSDTTFCEIILMEM	100
Qy	1881	ETHEDAWPFLPVLNKLVPYKVKIKKPMDFSTIREKSSQCPNLETFALDVLRLVFDNC	1940
Db	101	ESHDAWPFLPVLNKLVPYKVKIKKPMDFSTIREKSSQCPNLETFALDVLRLVFDNC	160
Qy	1941	ETFNEDDDSDIGRAGNNMRKYPEKWTDTFK	1970
Db	161	QTFNEDDDSEVGKAGHMRPFESRWFYQ	190
RESULT 26			
US-09-698-295-10			
; Sequence 10, Application US/09698295			
; Patent No. 6689584			
; GENERAL INFORMATION:			
; APPLICANT: Jones, Michael H.			
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATORY FACTOR			
; FILE REFERENCE: 06501-068001			
; CURRENT APPLICATION NUMBER: US/09/698,295			
; PRIOR FILING DATE: 2000-10-27			
; PRIOR APPLICATION NUMBER: PCT/JP99/02340			
; PRIOR FILING DATE: 1999-04-30			
; PRIOR APPLICATION NUMBER: JAPAN 10/137631			
; PRIOR FILING DATE: 1998-04-30			
; NUMBER OF SEQ ID NOS: 10			
; SOFTWARE: FASTSEQ for Windows Version 4.0			
; SEQ ID NO 10			
; LENGTH: 2781			
; TYPE: PRT			
; ORGANISM: Homo sapiens			
US-09-698-295-10			
Query Match 3.8%; Score 390; DB 4; Length 2781;			
Best Local Similarity 18.1%; Pred. No. 3.8e-18; Indels 948; Gaps 105;			
Matches 477; Conservative 322; Mismatches 890;			
Qy	17	NOEQSKNP-LDARVDKIKKKPRKXAMESSNSDSSTSGTSSDTSSEGISSSDDLEED	75
Db	400	NKARGSNKSPFLAANEELLESIRAKG--DIDNVKSPETEKKQNE--TENDSKDAEKN	454
Qy	76	EHE-EDQIERSEDDSDSESEAHKNNQVLHGIDPKADQKQAKTEKAEK-	127
Db	455	REEPEDQSLKSDSDKTPDDPEQKSE---VGDFKSEKSNSELSPGAGKAGSSTR	510
Qy	128	--RIHQPLPLAFESQTHSFQSQKQPVLSQQLPFIFQSSQAKES---VNKHTSVIQ	180
Db	511	ITIRLNP-----DSKLSQKSSQV-----AAAAHAENKLFKEKEVLVNV	551
Qy	181	STGLSVNPKLSLVNQAKKTYMKLI1VSPDVLKAGNKNT-----SEE	223
Db	552	SGGEISRL-----STKKEVIMK-----GNINNYFKLGQSKYRVVHNOYSTN	593
Qy	224	SLLTS-----ELRSKREYKQAF---PSQLKQSSSKLKKVIAALSNPATSSSPAHKP	276
Db	594	SPALNKHQHREDHDKRRHLAHKFLCTPAGEFKWNGSVHGSVKVL--TISTRLTITQ	647
Qy	277	QTLNHNPNPLTALLGNHQPNGV---IQSVLOEAPLALTTKMKQSKINENIAAAS	332
Db	648	--LENNIPSSFL-HPNWASHRANWIKAVQMCSPREFALAI-----LECAVK	693
Qy	333	TPFSSPVLNSTSGRRTPGNQTPVMPSPASPLHSGGKEKAVSNVNPVKTQHHSHPAKSLV	392
Db	694	PVWMLPIWREPLGHLRLHRTSIEEREKEVKKEKEQEEETMQQATWTKYTFPVKHQV	753
Qy	393	-----EQPRGT-----DSDIPSSKQSDSNEDEEED--	423
Db	754	WKQKEEVRVTCYGGWSWISKTHVYRVFVKLPNGTNNVYKSLSEGTKNNDENDESDDR	813
Qy	424	-----EEDDEDE--SDSQSDSNSSDTSGSEEDDDDDQDSDSDSDEGE	470
Db	814	KCSRSPKKIKTEPSEKDEKVGSDAAKGAQDNEMDIKITEKKDQDVKLELSDSDSKPK	873


```
Db 2281 TLQGENKELLSOLETRHLYHSSQNELAKLESELKSLKDQLTDLNSLEKCKEQKGNLEG 2340
Qy 1658 -----EDSSAHALERKSDNPLDIAVTRLADLE--RNRIERIBEDIAPGLRWRRALS 1708
Db 2341 IIRQOEADIQNSKFSVEQLE-----TDLQASRELTSRLHEEINMKEQKIIISLS 2389
Qy 1709 EARSAAQVALC--IQOLKSIAWEKSIMKYVQICRKGDNBEILLLLDCGCKGCHTYCHR 1766
Db 2390 GKEAIAQVAIAELROOHDKEI---KELENLLSQ-----EEENIVLEENKK----- 2433
Qy 1767 PRITTPDGDWPCACIASG--OTLK-IKKLHVKGK-----KTNESKKGKVTLTG 1816
Db 2434 -----AVKTNQLMETLTKIKENTQQAQLDSFKVSMSSQLNDRDRIVG 2478
Qy 1817 DTED-----EDASTSSSLK--RG----- 1833
Db 2479 DYQOLEERHLSILEKQDQIQEAAAENKKEIRGLRSHMDLNSNAKDAELIQYRE 2538
Qy 1834 -----NKDLQKR-----KMEENTSINLSQESFTSVKKPKR 1864
Db 2539 DLNQVITIKDSQKOLLEVQLOONKELENKYAKLEBKLESEBEANEDLRRSFNALQEBKQ 2598
Qy 1865 DSDKDALCSMLITEM 1880
Db 2599 DLSKEIESLKVSIQSOL 2614

RESULT 28
US-09-949-016-11433
; Sequence 11433, Application US/09949016
; Patent No. 6812339
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11433
; LENGTH: 2733
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-11433

Query Match 3.7%; Score 372.5; DB 4; Length 2733;
Best Local Similarity 18.5%; Pred. No. 6.5e-17;
Matches 402; Conservative 327; Mismatches 779; Indels 665; Gaps 84;

Qy 19 EQSKNQLDARDYDKDKKPR--KKAMESSSSSDSGTSSDTSSEGISSSDDLEBDE 76
Db 270 EQHLSLSIEA---KSKDVKIEVLQNELDDVQLFSEQSTLRSLOSQONKESEVLEGAE 326
Qy 77 EEDQSIEEEDDDSDSESAQHSKNNQVLLHGIDPPKADGOKATEKAQEKRIHQPLPLA 136
Db 327 RVRHSSKVEELSQALSKELEITKMDQLLLEKRDVETLQTIIEKQD---QVTEIS 381
Qy 137 FESQTHSFQSQK-----QPOVLSQLPFIPOSSQAKESVNYKITSVIQSTGLVSNKVP 190
Db 382 FSMTEKVMQNLNEKFSGLGVEIKTLKEQLNLLSRAEAKKEQVEEDNEV--SSGLQNYDE 439
Qy 191 LSLVNQAKETMYKMLVSPDVLKAGN-----KNTSESSLLTSELRSKRE 236
Db 440 MSPAQCIKSEELQHEP----DLTKKENEQRKKQLQAALINRKELLQVRSLEELANLKD 495
Qy 237 QYKQAFP---SQLKQESSKSLKVIAALSNPKATSSS-----PAHPKQTL-----ENNHP 284
```

```
Db 496 ESKKEIPLSETERGEVEEDKENKEY-----SEKCVTSKQCEIYIYKLOTISEKEVELOHI 550
Qy 285 NPFLTALLGNHQPNGVIOQAPALATTTKTK-----MQSKINEN----- 326
Db 551 RKDLEELAAEEQ---FQALVKQMNQTLQDKTQIDLLQAEISENQAIQKLIITSNTDA 606
Qy 327 -----IAAASPTSPSPVNLSTSGRRTPGNQTPVMPSPASPILHSGOGKAKVSNVNNVPKVT 381
Db 607 SDGDSVALVKETVISPCTGSSEHWKPELE-----EKILALEKEKEQKQKLEALT 659
Qy 382 QHSHPAKSLVEQRGTDSDIPSSKDSSEDSNEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDE 431
Db 660 SRKA-ILKKAKEKERHURLELKQKQD--DYNRLOEQFDEQSKENENIGDQLRQLOIQVRE 716
Qy 432 SDDSOSESDSNSE--SDTEGSEE-----EDDDKQDQESD----- 464
Db 717 SIDGKLPDQOESCSSTPGLEEPLFKATEQHHTQPVLESNLCPDMPHSEDASALOGGT 776
Qy 465 -----SDTEGKT--SMKLNKTTSSVKSPS---MSLTGHSTPRNLHIAKAPGSAPA 510
Db 777 SVAQIKAQLEAEKVELELVKSVSTTSSELTCKSEVQLOEQINKQGLEITESLKTVSHE 836
Qy 511 ALCSQSPAFILCTSSSTLTSSPHSGTSKRRVTDRELRIPILEYGHWQRETRIRNFGGL 570
Db 837 AEVHAESLQKLESSQLQIAGLEHL-----RELQPKLD-----EL 871
Qy 571 QGEVAYYAPCGKLRQYPEVIKYLSRNGIMDISRNPFSFAKIRVGDVEARDGQEQMOW 630
Db 872 Q-----KLISKKEEDVSYLS-----GQISEKAALTKQT 901
Qy 631 CLLKEEDVIPRIRAMEGRGRPPNPDQRQAREEESRMRKRGRPPNVGNAEFLDNADAKLL 690
Db 902 EIIQEEDLIKALHTQLEMQAK--EHDRIKQLOVELCEMKQKPEBIEGEE---SRAKQIQI 956
Qy 691 RKLQAEIARQAAQIKLRLKQKQEQ-----ARVAKEAKQQAQMAAEKKEKQEQIKI 744
Db 957 RKLQAALISRKEA--LKENKSQBELSLARGTIERLTLSLADVESQVSAQNKEKOTVLRGL 1015
Qy 745 -MKQEQKIKRIQIQRMEKELRAQOIL-----EAKKKKKEEAANAKLL 785
Db 1016 ALLQEERDKLITE--MDRSLLENQSLSSCSLKLALGLTEDEKLVKTESLSSKIA 1073
Qy 786 EA-EKRIKEKEMRRQAVLLKHQE-----RERRRQHMM-----LMKAMBARAKA 828
Db 1074 ESTEWQEKHELOXEYEYELLQSYENVSNAEIRIQHVVEAVRQEKQELYGKLRSTEANKKE 1133
Qy 829 BEKERLQKQEKDEKRLNKERKL-----EQRRLLEMAKELKKNEDMCLADQKPLPELPRI 884
Db 1134 TEKQLEAEQEMEMKEMKRRKFAKSQKQKILELE-----EENDRLRAEVHP----- 1179
Qy 885 PGLVLSGSTFSDCLMWVQFLNFGVLGFDVNDVPNLSVLQEGLLNIGDSMG-EVQDILL 943
Db 1180 -----AGDTAKECMTETLSSNASMKBEELERVMEYETLUSKKFQSLMSSEKDSUSEEVQDU- 1233
Qy 944 VRLLSAAVCDPGLITGYKAKTALGEHLNVLNVDNVSSEILQIFMEARHCGGTTELTESTUKT 1003
Db 1234 -----KHQIEGNVSRQANLEA---TEKHDNQTNVTEE-GT 1264
Qy 1004 KAPOHTPAQKASVLAFLINELACKSVSVSEIDKN-----IDYMSNLRKDKVNVVEGKRLK 1059
Db 1265 QSIPOETEOQSLSMS---TRPTCSESVPSSAKSANPAVSQKDFSSHDEINNVLLQIQDLQKE 1321
Qy 1060 RIHAKTKGKRTSGGIDLGEQHPGLGTPTPGRKERRKGGSDYDDDDDDDDDDDDDDDDDDDD 1119
Db 1322 RIAGLEBEKQKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKXKX 1376
Qy 1120 ED-----BEDKEDQKGGKTDICEDEDEGQAAASVBELEKQIEKLSKQSQYRRK 1168
Db 1377 EELSRVTKLETAEEKKDLE-----ERLMNQLAELNGSIGNYCQD 1417
Qy 1169 LFDA---SHSLRSVNFPGDPRYRRYRWILPRCGGIFVEGMEGEGLEETAKREKUKKAAES 1225
```

```
Db 1418 VTDAQIKNELLESEMKN-----LKKC-----VSELE---EEKQQLVKEKTKV---ES 1458
Qy 1226 VOIKEEMFETSGSLNCSNTHC-FQKEDLKEKONTNLFLOKP-----GSFSKLSKL 1276
Db 1459 EIRKEYLSKIOGAQKEPKGNKSHAKELQELLKEQEVKOLQKDCIRYQEKISALERTVKA 1518
Qy 1277 LEVAKMPPSESVMTPKPNAGANGCTLSYQNSGKSHSLGS--VQSTATQSNVEKADSNLNFN 1334
Db 1519 LEFVTEQSKOLEITKENLAQ---AVEHRKKAQAEALASFVKVLLDDTQSEAAVLADNL-- 1573
Qy 1335 TGSSGPGKFPYSLPNQLLTKTEKRNQWFSLLPRTPCDDTSLTHADMTASLVTPQSQP 1394
Db 1574 -----XLKKELOKESVK-----SQMKOKDELERLEQAEK 1607
Qy 1395 PSKSPPTAPLGSSAQNPNGLNPALSPLOV---KGGVSMGLQFCWPTGVVTSNTPF 1451
Db 1608 HLKEKNKMOEKLDALURREKVHJEE-TIGEIQVTLNKKDKVEQQLQ----- 1651
Qy 1452 TILSVPSLGLSLSGNNGNSFLTSNVASSKESPVQNEKATSAQAAAVEVAKPVDPPSP 1511
Db 1652 -----ENLDSVTQLAAFTKSMSSLODDR----- 1676
Qy 1512 KPPEMFGWRIIDPDALKALLKVLHRLGIREKALQKQIKHLDYITQACLKNKQVAI 1571
Db 1677 -----RVID-----EAKKWERKTSFSD-----AIQSKER-- 1698
Qy 1572 IELNENEEN-QVTRDIVENWSVEEQAMEMDLSVLOQVEDLERRVASASLOVKGWMC----- 1626
Db 1699 -EIRUKENCVSLKQDLQROMSIHMBELKINISRLSHDKQIWEKQAQTEVQLQKVCDTLQ 1757
Qy 1627 ---PEPASEREDLVVFEHKSFTKL-----CKEHDGEFTG--- 1657
Db 1758 GENKELLQLEBETHRLHYSSQNELAKLESELKSLKQDLTDLNSLEKCKEQKNLEGIIR 1817
Qy 1658 ----EDESSAHALERKSNPLDIAVTRLADLE--RNIERRIEEDIAPGLRVWRALSPEAR 1711
Db 1818 QOEADIONSKEFSYQLE-----TDLQASRELTSRLHEBINMKEQKIISLSCKE 1866
Qy 1712 SAAQVAGC--TQOLOKSTAWKSIWKVYCOICRKGDNELLLLDCGCKGCHTYCHRPKI 1769
Db 1867 EAIQVIAIAELRQOHDKET---KELENLLSQ-----EEENIVLEENK-- 1907
Qy 1770 TTIPGDWFCPACTAKAGS--QTLK- IKKLHVKGK-----KTNESKKGKVTLTGDTG 1819
Db 1908 -----AVDKTNQLMETLTKIKENIQKAQLDSFVKSMSSLQNDRDRIVGDIYQ 1955
Qy 1820 D-----BDSASTSSSLK---RG----- 1833
Db 1956 QLEERHLSIILEKDQLIQEAANAENKKEEIRGLRSHMDLNLSENKLDABLQVREDLN 2015
Qy 1834 -----NKLOQR-----KMEENTSINLSQESFTSVKKPKRDS 1867
Db 2016 QVITIKSQKQLEVLQONKELENYAKVLEKLKESFEANEDLRRSFNALQEBKQDLS 2075
Qy 1868 KDALCSMLTTEM 1880
Db 2076 KEIESLKVISQL 2088
```

RESULT 29

```
US-09-698-295-1
; Sequence 1, Application US/09698295
; Patent No. 6689584
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATORY FACTOR
; FILE REFERENCE: 06501-068001
; CURRENT APPLICATION NUMBER: US/09/698,295
; CURRENT FILING DATE: 2000-10-27
; PRIOR APPLICATION NUMBER: PCT/JP99/02340
; PRIOR FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: JAPAN 10/137631
; PRIOR FILING DATE: 1998-04-30
```

```
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 2907
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-698-295-1

Query Match          3.6%; Score 368; DB 4; Length 2907;
Best Local Similarity 17.9%; Pred. No. 1.5e-16;
Matches 488; Conservative 325; Mismatches 916; Indels 994; Gaps 106;

Qy 17 NOESKQNP-LDARVDKIDKPKKXAMSSSDSGTSSDTSSESIGSSDSDLDLED 75
Db 400 NKARGSNKFLAANEELLESIRAKG--DIDNVKSPETEEDKNE---TENDSKDAKN 454
Qy 76 EER-EDOSIERSEDDSDSSESAQHS-----NNQVLLH----- 108
Db 455 REEFEDQSLKDSDDKTPDDPEQCKSEEPTEVGDGKGNVSANLGDNTTNTATSETPSE 514
Qy 109 ----GISDPKADQKATEKAQEKRIHQPLP---LAFESQTHSFQSQOKQPOVLSQLP 159
Db 515 GRSPVGCLETPDSSNMAEKVASELPQDVPEEPNKTCESSNTSATTTSIQNLSENSS 574
Qy 160 FIFOSSQA-----KEESVNVKHTSVIOSTGLVSNV----- 188
Db 575 SELNSSQESAKAADDPENGERSHTPVSIQEEIVGDFTSR-KSTGELSESPGACKGASG 633
Qy 189 ----RPLSLVNOAK-----KETYMKLIVPS-----PDVLK 214
Db 634 STRIITRLNPDKSLQSKQVAAAHANKLFKEGKVLVNVNSQGEISRLSTKKEVIM 693
Qy 215 AGNKVT-----SESSLLTS-----ELRSKREYKQAF---PSQLKKQE 250
Db 694 KGNINNYFKLQGEQKYRVYHNYSTNSFALNKHQHREDHDKRRLAHKFCITPAGEFKWN 753
Qy 251 SSKSLKVIAALSNPATSSSPAHPKQTLNHNHPFLTNALLGNHOPNGV---IQSVI 306
Db 754 GSVGSKVL-TISTRLITITQ-----LENNISSFH-HPNWASHRANWIKAVQMSKP 804
Qy 307 QEAPLALTTTKMOSKINENIAAASSTFFSPVNLSTSGRTPGNQTPVMPSPILHSQ 366
Db 805 REFALALAI-----LECAVKPVMPLIWFREFLGHLRLHMTSIEREKEKVKKK 853
Qy 367 GKEKAVNNVNPVKTQHHSHPAKSLV-----EQRGT-----D 399
Db 854 EKQEEETMOQATWVKYTFPVGHQVWKQKGEEYRTVGYGWSWISKTHYRVFPKPLGN 913
Qy 400 SDIPSSKDSNDEDEDEDEED-----EEDDEDEDE--SDDSQSSEDSNSE 444
Db 914 TNNYRKSLEGTKNMNDENDESDKRCSRPKKIKIEPSEKDEKVGKSDAAKQDNEM 973
Qy 445 SDTEGSEEDDDDDKQDSESDSDTEGKTSMKLN---KTTSSVKSPSMLTGHSTPRNLHI 501
Db 974 DISKITEKQDVKELLSDSDKPCKEPMEVDMDKMTESHVN----- 1016
Qy 502 AKAPGSAALCSESQ-----SPAF-LGTSSSTLTSSPHSGTSKRRVTDRELRPL 553
Db 1017 -----QESSQVDVVNVNVEGFHLRTSYKKTKS-----SKLDGLLERRIKQFTL 1060
Qy 554 EYGMQRETRIRNFGRLQGEVAYVAPCGKLRQYPEVIKYLRSNGIM-DISRDNFSFAK 612
Db 1061 EEK-ORLEKIKLEGG-IKGIGKTSNNSKNLSSEPVITK--AKGGCQSDSNRQSQSPNAN 1116
Qy 613 IRVGDYFEARDGPOEM-QWCLLKEEDVI-----PRIAMEGRGRPP----- 653
Db 1117 -----NDQPEDLIQGCQSQSDSVLRMSDPSHTTNKLYPKORVLDVDSIRSPTKCP 1167
Qy 654 --NPDORAREESMRRRKGRPP-----NVGNAEFLDNADAKLLRKLQAEIARQAAQIKL 707
Db 1168 KQNSIENDIEEKVSDLASRQGEPTKSKTKGNDFFID--DSKL---ASADDITGLICKNK- 1221
Qy 708 LRKLQKQEQARVAKAKK--QQAIIAAABEKR----- 736
```


COUNTRY: U.S.
 ZIP: 10036
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/714,741
 FILING DATE: 16-SEP-1996
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Frommer Esq., William S.
 REGISTRATION NUMBER: 25,506
 REFERENCE/DOCKET NUMBER: 454312-2460
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (212) 840-3333
 TELEFAX: (212) 840-0712
 INFORMATION FOR SEQ ID NO: 32:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 8991 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: amino acid
 US-08-714-741-32

Query Match 3.6%; Score 362.5; DB 4; Length 8991;
 Best Local Similarity 17.4%; Pred. No. 2e+15;
 Matches 398; Conservative 338; Mismatches 763; Indels 787; Gaps 87;

QY 5 KSTSGGGR-----KCNQ--QSKNQPLDARVDKIDKKPKKAMES 45
 DB 1318 KATESGLVTRVEAEKNTDARQKLVLCNEVWLOAXXAELESGHKLPEPNKKMILT 1377
 QY 46 SSNSDSGDT-----SDTSSEGISSDDLDEDEEEEDQSIIES----- 86
 DB 1378 SLASKAIIAGLIVASQPTVRAEEAPVASQSKAEKDYDAARDAENAKKALEEAKRAQKX 1437
 QY 87 -EDDSDSE-----SEAQHSN--NQVLLHGIDPKADGQKATEKAQEKIHQPLP 134
 DB 1438 XEDDQKTEEKAKXDXQAEQANLXYQLLKQYVS-ESDGKKKKE----- 1484
 QY 135 LAFESQTHSFOSQKOPVLSQLPFIFQSSQAKESVKNKTSV-IQSTGLVSNVKPLSL 193
 DB 1485 -----XEXXADAAKKIEKXADLXIXQENKMKMILTSLASVAILGALVAS----- 1533
 QY 194 VNQAKKETMYKLIIVSPDVLKAGNKNKNTSESSLTSELRSKREQYKQAPPSQLKKQESSK 253
 DB 1534 -----QPTVRAEEAPVASQSKAEKDYDAAEKSKAAEEDLEEAQAQRKYDEDQK 1584
 QY 254 SLKKVIAALSNPKAT-SSSPHPKQTLNNHPNPLTNALLGNHPNPGVIOSVIOEAPLA 312
 DB 1585 KSEE-----NEKETEEASERQQAATLYHLESKEFLNYFQDNHRN----- 1625
 QY 313 LTTTKMQSKINENIAAASSTFFSPVNLSTSGRTPCNQTPVMPASPIILHSQKKEAV 372
 DB 1626 ---KKMILTSLASVAILGALVASP-----PTVRAEEAPVA-SQSK----- 1664
 QY 373 SNNVNPVKTQHHSHPAKSLVEQFRGTDSDIPSSKSDSDEDEDEDEDEDEDEDE 432
 DB 1665 -----AEKDYDTAKDAENAKKALEEAKRAQEKY 1693
 QY 433 DDSQSDNSGESDTEGSEEDDDDKO-----QDESDDSTEGER 471
 DB 1694 ADYORRIEBAKAKETHASLEQOEANKDYQLKKYLDGRNLSNSVLKEMEEAEKKDKE 1753
 QY 472 TSMKLNKTTSSVKPSMNLGTHSPNLIHIAKPGSAPAALCSQSQSPAFLGTSSTLTS 531
 DB 1754 KPAEFNKIRREIVPN-----POLEMAR-----RKSEVAK 1784
 QY 532 SPHSQTSKR-----RRVYDER-----ELRIPLEYGWQRETRIRNFGRLQGEVAYY 577

DB 1785 TKESGLVKRVEEAEEKVTEARPKLDAERAKEVILQAQIAMNKKKMILT-----SLASV 1837
 QY 578 APCGKKLRQYPEVIKYLNRNGIMDISDNFSFSAKIRVGDFYEARDGPOEMQWCLLKEED 637
 DB 1838 AILGAGLVASPTTVRAEEAPVASQSKAEKDYDTAKR--DAENAKKALEEAKRAQEKYAD 1895
 QY 638 VIPIRAMEGRGRPPNPDRQRAEESRMRKR--GRPPNVGNAEFL-----DNADAK-- 688
 DB 1996 YORRIEBAKAKETH-ASLEQOEANKDYQLKKYLDGR--NLSNSVLKEMEEAEKKDK 1952
 QY 689 -----LLRKLOAQEIAQAAQIKLL-----RKLOKQEOARVAKAKKQQAAMAA--- 732
 DB 1953 EKQAGLANK--KMILTSLASVAILGALVTSQPTLVRAEESPVASQSKAEKDYDAKRD 2010
 QY 733 -----BEKRKQEQIKIMQ--QEKIKRIQIRMEK-----ELRAQOILEAKKK 775
 DB 2011 AENAKKALEEAKRAQEKYADYORRIEBAKAEQQAQLEQOEANKDYQLKKYLDGRNLS 2070
 QY 776 KEEAANAKLLEAEKRIEKEKMR-----RQQAALLKHQERERRRRQHMLMKAMEAR--KKA 829
 DB 2071 NSSVLKEMEEAEKKDEKQAEFNKIRREIVVPNPQLEEMARRKSEVVKAKESGLVKEVE 2130
 QY 830 EKERLQEKRDEKLNKERKLEQRRLLEMAKELKPNEDMCLADQKPLPELPRIPLGLVL 889
 DB 2131 EAEKKVTEAR-----QKLDARAKE----- 2150
 QY 890 SGSTFSDCLMVVQFLRNFQKVLGFDVNIQVNLVLOEGLLNIQSDSMEVODLLVRLISA 949
 DB 2151 -----VLQPTR-----VENEVHKLQKMNKKMILTSLASV 2182
 QY 950 AVCDPGLITGYKA-----KTALGHELLNVGNRDNVSEILQIF-MEAAHCGOTE 996
 DB 2183 AILGAGLVTSQPTVRAEESPVVVEKSSLEKYEAKADTAKKDYETACKKAEDAQKK 2242
 QY 997 LTESLK--TKAFOAHTPAQASVLAFLI-----NELACSKS-----VSEID 1036
 DB 2243 YEDDQKTEEKARKEAEASQKLDVALVQWVAYKEYREVQNRQSKYSDADYQKLTEDV 2302
 QY 1037 KNIDYMSNLBRDKWVVEGKLRLIHA-----KTKGRD----- 1071
 DB 2303 SKTB---KARKEODLQNNFNNEVAVVADPTCVGDXRMNKKKMILTSLASVAILGAGX 2359
 QY 1072 -TSGGIDLGSEHQPLGTPGKRKRKGGSDYDDDDDDDDDDQDQDDEDEDEDEDEDEDE 1130
 DB 2360 VTSQPTVRAEEAPVVEKSSLEKYEAKADYDAKKDYDEAKKAEAKKYEEDQ-- 2417
 QY 1131 KKTDCEDDEGDOAASVEELEKQIEKLSQOOSYRKLFDASHLSRSMVPGPDYRRRY 1190
 DB 2418 KKT---BEKAEKAKAAS-ESIAKATEEVQKAVLDYITAIRNHDS----- 2458
 QY 1191 WILPRCGGI FVEGMESEGLEETIAKEREKL---KKAESVQI-----KEEMFE 1234
 DB 2459 -----GKTSAAEAENKAKERDYCCAGKKFDP1QTPFVASITQMTLMNKKKMIL 2506
 QY 1235 TS-----GDSLNCNTDHCQEKEDLKEKONTNLFLQKPGSFSKLS 1274
 DB 2507 TSLASVAILGALVASPTTVRAEEAPVASQSKAEKDYDTAKRDAEN-----AK 2555
 QY 1275 KLLVAXKMPPESEVMTPKPNAGANGCTLSYONGSKHSLGVSQSTATQSNVEKADSNLNF 1334
 DB 2556 KALBEAKRAQEKYA-----DYQ-----RRIEEKAAKETQASLEQOEANKDYQ 2597
 QY 1335 TGSQGPCKFYSPNDOLLKLTLEKNQWPSLLPRTPCDDTSLTHADMSTASLVTPOSQ 1394
 DB 2598 L-----KLKYLQGRN-----LSNSVLKEMEE 2621
 QY 1395 PSKSPSPAPLPGSSAQNVPGLNPFALSPLOVKGVSMMGLQFCGWPTGVVTSNIPFTLS 1454
 DB 2622 AEKKDKENQAEFNKIRREIVVPNPQLEEMANMKKMILTSL-----AS 2664
 QY 1455 VPSLGSGLHSEGNNGNSFLTSTNVASSKS--ESPVPQNEKATSAQPAAVEAKVPFPSPKP 1513
 DB 2665 VAILGAG-----FVASQPTVRAEESPVASQSKAEKDYDAKKDAK-----NAKK 2709


```
Db 997 GQKKSIKRRKSKGSDSDISDEE-----AAQSKKKRG----- 1030
Qy 1187 RRRYILPRCGI-----FVEGMESGLEETAKEREKKAESVOIKEEMFETSGDSL 1241
Db 1031 -----PRTPPTTKBELVEMCNKGKNGILEDQSKKE-----DTAFSDMSD--- 1069
Qy 1242 CSNTDHCQKEDLKEKDNNTLFLQKPGSF-SKLKLLVAKMPPSEVMTKPNAGANGC 1300
Db 1070 ----EDVDRTEVTAETH-ATATTPGSTPFLSLL-----PP-----PPVNTATAT 1113
Qy 1301 TL-----SYQNGKHSLGSVOSTATQSNVEKADSNNTFNTGSSGPKFYSPLPNDQ 1351
Db 1114 TVPATLAATAAATSAITISTSATPTNT-----TNNTF----- 1151
Qy 1352 LLKTLTEKRWQFSLLPRTPCDDTSLTHADMSLASLVTPOQPPSKSPSPAPLGSQAQ 1411
Db 1152 -----ANEDSHRKCHRTVRKEVETPHVTIEDAQHRKPMQKRS-----LGSNRS 1197
Qy 1412 NPVGLNPFALPLQ---VKGVSMGM---LOFCGWPTGVVTSNIPFTLSVPSLGSGLGLS 1465
Db 1198 NRSHTSGRLRSPNSAHRSGDQGRKRVLHSGSRDREKTKSLITGERKSRIDQLKRG 1257
Qy 1466 EGNNSFLTSNVASKSPVPQNEKATSAQAAVEVAKPV-----DPPSPKPIPEE--- 1517
Db 1258 E-----PSRSTSDQDSRSHSRSSPSDRQVHSRSGSFDSDRLQERDRY 1305
Qy 1518 -----WQFGW-----W-----RIIDPEDLKALKVLHLRGIREKALOK 1550
Db 1306 EHDRERERERDTRQWDRDADKQWPNRDRDLRERERERERDRDLRERERERLSD 1365
Qy 1551 QIQKHLDTQACLKNKDVAIELNENE-ENQVTRDIV-----ENWSVEEQAWMDLSVLQ 1605
Db 1366 SVDRDRDRDRBTFSQIESVKRCEAKLGEHERDLSTSRDLSALDKERMDKDLGSVQ 1425
Qy 1606 QVEDLERRVASLQVKGMMCEPAPEREDLVYFEHKSFTKLCKEHDGFTGEDES---S 1662
Db 1426 GFEDTNKSTESLE-----AGDDDES KLD 1450
Qy 1663 AHAL-----ERKSNPLDIATVRLADLERNIERRIEREEDIAPLRV 1702
Db 1451 AHSLSGAGEGYEPISDDELDEILA--GDAEKREDQDQDEKMPDPLDV 1496

RESULT 32
US-10-144-198-4
; Sequence 4, Application US/10144198
; Patent No. 6833247
; GENERAL INFORMATION:
; APPLICANT: Origene Technologies Inc
; TITLE OF INVENTION: Regulated Prostate Cance Genes
; FILE REFERENCE: 9U 105 R1
; CURRENT APPLICATION NUMBER: US/10/144,198
; CURRENT FILING DATE: 2002-05-14
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 1564
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-144-198-4
```

```
Query Match 3.5%; Score 357; DB 4; Length 1564;
Best Local Similarity 18.5%; Pred.No.3.5e-16;
Matches 342; Conservative 255; Mismatches 683; Indels 568; Gaps 70;

Qy 8 SSGGNRKNQKQSKQNPDLARVDKIKDKPRKXAMSSSSDSGTSSTSGTSSG 67
Db 64 SPRKGYSNYSRSPERTGDLREMKRQDVTEPQKRTTBSSSPVKESRG-RHR 122
Qy 68 DSDDL-----ESEDDEEDOSIE-ESEDDSDSESAHQHKNNOVLHGISDPKADQKATE 122
Db 123 EKEDIKIKTPTPSEENVEWETNRDSDN-----GDINYD----- 159
```

```
Qy 123 KAOEKRIHQPLAFESQTHSFQSQOKOPOVLSQLPFIFQSSQAKESVNKHTSVIQST 182
Db 160 -----YVHLSLEMKRQKIQRELM-----KLEQENMEKREIIL--- 192
Qy 183 GLVSNVPLSLAVNOAKKETYMKLIVSPDVLKAGNKNTSESSLLTBELRSKRQYKOAF 242
Db 193 -----IKKEVSPVVRSKLSPSPS-LRKSSKSPKRKSSPKSSS-ASKDKRTSAV 240
Qy 243 PSQLKQBE-----SKSILKKVIAALSNPKATSSSPAHPKQ-TLENNHNPFTNALLGNHQ 297
Db 241 SSPLLDQQRNSKTNQSKKK-----GPRTPSPPIPEDIALGKKYKVKVDRIEKT 294
Qy 298 PNG-----VIOSVIOEAPLALTTKTMOSKINENIAAASSTPFSVPVNLSTGRR--- 347
Db 295 RDGKRGDRFERQREKDKFRSTSPAGQHHSPIISRHHSSSSQSGSIIQRHSPFRKRRT 354
Qy 348 -TPGNQTPVMP-----SASPI-----LHSQGREKAVSNVNVVKTQ-HHSHPAKSILVEQPRG 397
Db 355 PPSYQRTLTPLLRASAPVPSHLSPPQRKQSPRHRSPMREKGRHDHRTS----- 407
Qy 398 TDSIPSKDSESDNEDEED-----DEEDEDEDEDESDSDSESDESDEGSEE 452
Db 408 -----QSHRRRHERREDTRGDRDKDSREEREYEQDQSSDRDRDPRDGRDR 458
Qy 453 EDDDDKQDESDDTEGEKTSMLNKTTSSVKSPMSLITGHSTPRNLHIAKAPGSAAPAL 512
Db 459 RDADTRDRRELDRDRMRSREMRDYSRDTK----- 490
Qy 513 CSQSQAPFLGTSTLTSSPHSGTSKRRRVTDRELRIPLE-----YGMQR-----ETRI 563
Db 491 --ESRDP-----RDSRSTRDAHDYDRDEGRDTHRKEDTYPEESRSYGRNHLRESSRTEI 543
Qy 564 RNFG-----GRLOGEVAYVAPCGKQLRQYPEVIKYLSRNGIMDISRDNFSFS 610
Db 544 RNESRNSRSEIRNDRMGRSRGRVP-----ELPEKSGRSRGSQID-----SHS 587
Qy 611 AKIRVGDFYARQDGPQMWQCLLKEDVIPIRAMEGRRGPPPPDQRAEERMRMRK 670
Db 588 SNSNYHDSWETRSTYPE-----RDYPERDNRDQARDSFERH 626
Qy 671 GRPPNVGNAEFLDNADAKLRKLOAQIAQAAQIKLLRKLOEQEQARVAKEAKKQQAAM 730
Db 627 GERDRDRNRERDQRPSPPIRHQGRNDELRD-----ERREERRVDVDRDRDE-- 674
Qy 731 AAEKRRKQKQIKIMKQOEKIKRIQIIMEKELPAQOILEAKKKKKEAANAKLLEAB-- 788
Db 675 RARERDRERERERERERERERERERERERERERERERERERERERERERERERER 734
Qy 789 -----KRIKEBMRROQAVLLKHQERRRRRHHMLMKAMEARKKAEKERL-----KQ 836
Db 735 RERERERERERERERERERERERERERERERERERERERERERERERERERERER 794
Qy 837 EKRDKEKRLNKERKLEQRRL---ELEMALKKPNEDMCLADQKPLPPLPRIGLVLGSGT 893
Db 795 KGRDRREKREETREDRNPDRGHDERSKKYRNEGSPSPQSPKRREHSPD----- 847
Qy 894 FSDCLMVVQFLRPFKVLGFDVNDVNLVQLQGLLNI GDSMGCEVDLLVRLLSAAVCD 953
Db 848 -SDA-----YNSGDDKNKEH-----RLLSQVY-- 868
Qy 954 PGLITGYKANTALGEHLNVGNRDNVSEILQIFMEAHCGOTELTSLTKT-AFOAHTPA 1012
Db 869 -----RPQESRSL-----SPSHUTEDRQGRWKEDRKPE 897
Qy 1013 QKASVLAFLINELACSKSVVSEIDKNIDYMSNLRDRKWVE-GKLRKLRILIAKTKGRD 1071
Db 898 RKESSRRRYEEQEL---KEKVSSVDKQ-----REQTEILLESSRMRAQDIIHHQSEDE 947
Qy 1072 TSGGIDLGESQHPGLGTTTPGRKRRKCGDSDYDDDDDDSDQDQD-----EDDEDEDEKE 1126
Db 948 TS---DRAHDENKKAKIQQKPIKKK-----KEDDVIGIERGNIETTSDDGVFPKK 996
Qy 1127 DQKCKTDDICEDEDEGDAQASVELEKQIEKLKQSQOYRRKLPFDASHSLRSVMFGPDY 1186
```



```
; TYPE: PRT
; ORGANISM: Kaposi's sarcoma-associated herpesvirus
US-09-410-399-2

Query Match      3.2%; Score 326.5; DB 4; Length 1162;
Best Local Similarity 16.9%; Pred. No. 3.4e-14;
Matches 196; Conservative 191; Mismatches 405; Indels 371; Gaps 31;

QY 207 VPSDVLKAGNKTSESSLLTSELRSKREYKQAFPSQLKQESSKSLKKVIAALSHPK 266
DB 114 IPPSHVSPGTTDTHSPSPAL-PPTQSPESQRPLSLPTGRDSDSTMPRP-----PP 165
QY 267 ATSSSPAH-----PKQTLNNHPNFLTAL-----LGNHPNGVTSIQEAPLALT 314
DB 166 SOQTTPHSPPTTPPEPPSKSPDLSAPSTLSLRKRLSSPQGSTLNPICQPPV--- 222
QY 315 TKTKMQSKINENIAAASSTP---FSSPVNL-STSGRTPGNQTPVMP-----SASPILHSQ 366
DB 223 -----SPPRCDFANRSVYPWPATESPIYVGSSDGDTPPRQPTSPISIGSSSPSEGSW 276
QY 367 G-----KEKAVSNVNPVKTHHSH 387
DB 277 GDDTAMLVLLAIEAEASKNEKESNNQAGDNGDNEISKESQVDDKDNKODDEEQE 336
QY 388 AKSLVEQFRGTDSDIPSSKDSNDEDEDEDEDD-----EDDESDDSQSESNSN 442
DB 337 TD---EEDDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDE 393
QY 443 SESTEGSEEDDDKQODESDSDTEGKTSMKL-----NKTTSVKSPMSLTGHSTPR 497
DB 394 EEEDEEEDDDDDNEDNEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDE 452
QY 498 NLHIKAPGSAPALCSQSQAPLGTSSSLLTSSPHSGTSKRRVTRDERELRIPLEYGW 557
DB 453 -----QQQEPQ-----QQEPLQEQPQQEQPQQEQPQQEQPQQEQPQQEQP 486
QY 558 QRETRIRNFGRLQGEVAVYAPCGKRLQYPRVIVKYLNRNGIMDISRDNFSFSAKIRVGD 617
DB 487 QQEPQ-----QQE-----PLQEPQQEQP----- 505
QY 618 FYEARDGQEMQWCLLKEDVIPTRAMEGRGRPPNDRQ-AREESRMRRRKGRPPNV 676
DB 506 ---QEQEPQEQEPQEQEPQEQEPQEQEPQEQEPQEQEPQEQEPQEQEPQEQEPQEQ 561
QY 677 GNAEFLDNADAKLLKLAQEIARAQAQIKLLRKLQKQEQARVAKEAKQQAIAAEKR 736
DB 562 REPQ-----QREPQREPQREPQREPQREPQREPQREPQREPQREPQREPQREPQ 610
QY 737 KQKEQIKIMKQEKIKRIQIRMEKELRAQQTLEAK--KKKEEAANAALLAEKRIKEK 794
DB 611 QODEQOQDEQOQDEQOQDEQOQDEQOQDEQOQDEQOQDEQOQDEQOQDEQOQDEQO 670
QY 795 EMRRQQAALLKHQERRRRQHMLMKAMEARKKAEKELKQEKDEKRLNKERKLEQRR 854
DB 671 QODEQOQDEQOQDEQOQDEQOQDEQOQDEQOQDEQOQDEQOQDEQOQDEQOQDEQO 730
QY 855 LELEMAKELKPNEMCLADQKPLPRLPGLVLSGSTTFSDCLMVVQFLRNFVKVLGFD 914
DB 731 QODEQOQDEQOQDEQOQDEQOQDEQOQDEQOQDEQOQDEQOQDEQOQDEQOQDEQO 752
QY 915 VNIDVPNLSVLQGLNLNGDSNGEVQDILLVRLLSAAVCDPGLITGYKAKTALGEHLNVG 974
DB 753 -----QEQEQEQE----- 760
QY 975 VNDRNVSELIQIFMEACHQGTLETSLTKYAFQAHTPAQKASVLAFLINELACSKSVVSE 1034
DB 761 QOEEQEQELEEQEQELEEQEQELEEQEQELEEQEQELEEQEQELEEQEQELEEQEQELE 820
QY 1035 IDKNIDYNSNLRRDKVWVEGKRLKRIIHAKTGKRTSGGIDLGEBOHPLGTPTPGKR 1094
DB 821 QEQELEEQEQ-----ELQEQEQL----- 839
QY 1095 RRKGGSDYDDDDDDSDQDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDE 1154
```



```
QY 1439 -----GHTPTGV-----TSNTPFTLSVPSLGG---LGLSEGNNGNSFLTSTNVASKSES 1484
Db 1460 LVKEMQLGLEGLVPSLSSCPDSSSLSSGLDSSFYRALLQETGDMSSLLNLEGAVSAN 1519
QY 1485 PVPQNEKATSAQPAAVEAKVPDFSPK-----PIPEMQGWWR-----IID 1527
Db 1520 QCSVDEVFVSSLSQTYVDSLUKENLVSTNLRNFQDGLVKEMQLGLEBGLVPSLSSCPVD 1579
QY 1528 PEDLKALLKVLHLRGIREKALQKIQKHLDYITQACLKNKDVAIITELNENENQVTRDITV 1587
Db 1580 SSSLSLGDSSFYRALLEQTGDMSSLLNLEGVSA--NQCSDVDEVFVSSLQENLTRK-- 1635
QY 1588 ENWSVEEQAMENDLSVL-----QVEDLERRVASASLQVGMWCPPEPASEREDLVYFEHK 1642
Db 1636 ETPSAPAKGVE-ELESCLVEVYRQSLKELEKES-----QGTMKNKEIQELBQLLSERQ 1689
QY 1643 SFTKCKEHGDE-----FTGEDESSAHLERKSDN-PLDIATVRLADLERNIERR 1691
Db 1690 ELDCLRKQYLSENEQWQKLTSTVLEMESKLAEEKQTEQLSLEVARLQ----- 1740
QY 1692 IEEDTAPGLRVRRAL-----SEARSAAQVALCICQLOKSIWAKSINKVYQICRKGDN 1746
Db 1741 -----LQGLDLSRSLGIDTDAICQRNESCDSISKEHTSETTTPKHVDVHQICDKDAQ 1795
QY 1747 BELLLCDGCDKCHTYCHRP-----KITTPDGMFCPA-----CIAKAS----- 1787
Db 1796 QDLNL--DIEKITETGAVKPGCEGSPDYNPEPCEDKTQGSSECSISLSFSGPNA 1852
QY 1788 -----GQTLKIKLHVKGKTNESKGGKVTLLGDTDEDS--ASTSSSLKRGKNDIQ 1838
Db 1853 LVPMDFLGNQEDIHNLQVRKETS-----NENRLLHVIEDRDKVESLLENMKELDSKLH 1908
QY 1839 KRKMEENTSN--LSKQESFTSVKPKRDRDSKDALCSM---ILTEMETHE 1884
Db 1909 LOEVQMLTKIEACIELEKIVGLKKNSDLSEKLEYFSCDHOELLQVETSE 1960
```

```
RESULT 41
PCT-US95-16216-1
; Sequence 1, Application PC/TUS9516216
; GENERAL INFORMATION:
; APPLICANT: Yen, Timothy J.
; APPLICANT: Ratthner, Jerome B.
; TITLE OF INVENTION: Nucleic Acid Encoding a Transiently
; TITLE OF INVENTION: Expressed Kinetochores Protein, and Methods of Use
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dann, Dorfman, Herrell and Skillman
; STREET: 1601 Market Street
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/16216
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/353,700
; FILING DATE: 09-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Reed, Janet E.
; REGISTRATION NUMBER: 36,252
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 563-4100
; TELEFAX: (215) 563-4044
; INFORMATION FOR SEQ ID NO: 1:
```

```
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3248 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
PCT-US95-16216-1

Query Match
Best Local Similarity 3.1%; Score 316.5; DB 5; Length 3248;
Matches 414; Conservative 322; Mismatches 797; Indels 739; Gaps 91;

QY 22 KNQILDARVDIKDKPRKKAEMSSSDSGTSDTSSSEGISSDSDDLDEEEDQ 81
Db 19 KIQLEGLQDLKKEKQKQ-----FQDLSLEAAQKQKQ 53
QY 82 SIESEDDSDSESESAQHKSNNQVLLHGISDPKADGQKATEKAQKRIHQPLAPESOT 141
Db 54 KYENEK-----TEGTLNKRQRL-----M 73
QY 142 HSPQSOQKQPOVLQOOLPFIQSSQAQKESVNHKTSVIOSTGLVSNVPLSLVNOAKKET 201
Db 74 EICESLEKTKOKISHL-----QVRESQVN-----FOEQGLNSGKKQIEKLEQELKRC 121
QY 202 YMKL-----IVSPDVLKAGNKNTSESSLLTSEL-----SKREOVKQAPPSOLKQ 249
Db 122 KSELSRQQAQASADV-SLNPCNTPK--IFTPLTPSQYSGSKVEDLKEYN--KEV 175
QY 250 ESSKSLKVKVIAALSNPKATSSPAHPKQTLNHNPNPFLTNALLGNHQPNGVIOQA 309
Db 176 EERKLEAEVKAQAKAS-----QTL-----POATMNRHDIARHQASSSVFQWQ- 222
QY 310 PLATTKTKMQSKINENIAAASSFPFSPVNLSTSGRRTP-----GNQTPMPSA 359
Db 223 -----KTFPSHLLSSNSQRTPIRRDFSAFSGSELEVTPSR 256
QY 360 SPILHSQGEKAVSNVNPVTKQHHHPAKSLVEQFRGTDSIPSSKDSSEDEDEED 419
Db 257 STL--QIGKRDANSSFFGNSSPH-----LLDQLKAQNLKRNK-----INLELRQ 302
QY 420 EEEDEEDEDDESQSDSESNSSESDETEGSEEDDDDKQDE-----SDSTEGBK 471
Db 303 GHEKEMQGVNKFQELQLE---KAKVELIEKEKVLNKRDELVRTTAQYDQASTKYTA 359
QY 472 TSMKLNKTTSSVKSPMSLTGHSFTRNLHIAKAPGSAFALCS-----ESQSPAFIG 523
Db 360 LEQKLKLTEDL-----SCORQ-----NAESARCSLEQKIEKEKEFEELS 401
QY 524 TSSSTLTSSPHSGTSKRRRVTDRELRIPLEYGWQRETRIRNFGRLQGEVAYVAPCGKK 583
Db 402 RQQRSPQTLQDECIQMKARLTQELQ-----QAKMHNVLQAL-----DK 441
QY 584 LRQYPEVIKYLNRNGIMDISRDNFSFSAKIRVGDYFYEARDGQPMQWCLKEEDVIPRIR 643
Db 442 LTS-----VKQLENNLEB-----FKQLCRAB-----QAFQASQIKENELRRSME 482
QY 644 AMEGRGRPPNPDQRARE-----ESMRRRKGRPPNVGNAPFLDNADAK-----L 689
Db 483 EMKKNILLKSHSEQKAREVCHLEALNKIKQCLNQSN--FAEEMKAKNTSQETMLRDL 540
QY 690 LRKLQAEIARQAQAIKL-LRKLQKQEQ-----RVAKKAKQQAIA 731
Db 541 QEKINQNSLITLEKLAVALDEKQRDCSDLLKKRHHIBQLNDKLSKTEKESKALLS 600
QY 732 A-EKKRQKQEQIKMK-----QSEKIKRIQOIRMEKE-----LRAQOI----- 768
Db 601 ALELKKKEYELKEBKTLFSCWKSENEKLLTQMESEKENLQSKINHLETCLTKQIKSHE 660
QY 769 -----LEAKKKKKEEAANAKLFAEKRIKEKEMRROQAVLLKHQERRRQ 814
Db 661 YNERVRTLEMDRENLSVEIRNLHNVLDKSEVEVETQ-KLAYMELQQAEEFSDQKHQETE 719
```

```
Qy 815 HMMLKAMEARKAEERLKOEDKRLNKKERKLEORRLELWAKELK-----PN 867
Db 720 NWCL-KTSQLTQVEDLEH-KLQLLSNIMDKORCYQDLHAEYSLRDLKSKDASLVTN 777
Qy 868 ED-----MCLAQKP-----LPFLPRIPLGLVLSGSTFSDCLMVVQ-----FLRNF 907
Db 778 EDHQSLLAFQOQPAHMHSPANITGEQSGMP-----SERSECLREADQSPNAILQNR 831
Qy 908 GKVLGFDVNIIVPNLSVLQEGLLNTGDSMGVEQDQLLVRLLSAAVCDPGLITGYKAKTALG 967
Db 832 VDSLEFLESQKQMSDLQKCEELVQIKGBIEENLMK-----AEQ 872
Qy 968 EHLNVGNVRNVSEILQIFWEAHCCQTELTESLTKAFQAHTPAQKASVLAFLINELAC 1027
Db 873 MHQSFAETSQRISK-LOEDTSAH--QNVVAETLSAL-----ENKEKELQLLNDKVET 922
Qy 1028 SKSVSEIDKNIDYMSNLRROKVVVEGKRLRLRII-HAKKTGKRDTSGDILGEEQHPLG 1086
Db 923 EQAEIQE-----LKSNHLLSDLSKELQLLSETLSLEKKEMSIISLNKRETEL 972
Qy 1087 TPTPKRK-----RRKGSDDYDDDDDDDDDDOQD-----1116
Db 973 TOENGTLEINASLNQKXNLIQKSESPANVIDEREKSISELSQYKQKLLILQRCBT 1032
Qy 1117 -----EDDEDEKEDQKCK-----TDICEDEGDQAAASVEELEKQIEKLSKQSQ 1164
Db 1033 GNAYEDLSQKYAAQKQKSLKLECLNCTSLCENRK-----NELEQLKEAFKHEQ 1084
Qy 1165 YRRKLFDA-----SHSLRSMVFGPDYRRYRWILPRCGGIFVEGMESEGL 1210
Db 1085 FLTKLAFABERNQNLMLLETVOQLRSEMNDQNNSK-----SEAGGLKQEIIM- 1133
Qy 1211 BEIAKERELKKA--ESVQIKEMFETSGDSLNCSTNDHCBQKEDLKEKN-----1259
Db 1134 -TLKEEQNMQKQVNDLLQENELQMKVMKTKHEQNLSEPIRNSVKERESERQCNFKP 1192
Qy 1260 -TNLFQKPGFSKLSKLELVAKMPPESEVMTPEKNAGANGCTLSYQNSGK-----HSL 1312
Db 1193 QMDLEVKESILSDSYNAQLVLEAMLRNKE-----LKLQSEKEKECLQHEL 1238
Qy 1313 GSVOSTAQSNVEKADSNLNFQGS-----SGPGKFSVPLPND-----QLLKT 1356
Db 1239 QTIKRDLETSNLQMQSQEISGLKDCEDIDAEKISGPHL-STSQNDNAHLQSLQTTM 1297
Qy 1357 TEKMR-----QWFSLL-----PRTPC-----DTSLSLTH 1379
Db 1298 NKLNELEKICEILQAEKYELVTELNDRSRSECTATRKMAEVRGKLLNEVKILNDDSLGH 1357
Qy 1380 ADMSTASLVTPSQPPSPKSPPTAPLQSSAQNPVGLNPPFALSPLQVKGGVMMGL-----1435
Db 1358 GEL-----VEDIPGFEQGEQNEQHPVSLAPLDESNSYEHLTSLDKE 1399
Qy 1436 -----QFC-----1438
Db 1400 VQMFAPAELOEKFLSLQSEHKILHDQCMQSMSELOQTYVDSLKAKENLVLTNLRNFGD 1459
Qy 1439 -----GWPTGV-----TSNIPFTLSVPSLSG-----LGLSEGNCSFLTNSVASSKES 1484
Db 1460 LVKEMQLGLEGLVPSLSSSCVPDSSSLSLGDSFYRALLEQTQDMSLLSNLEGAVSAN 1519
Qy 1485 PVPQNEKATSAQPAAEVAKVPDFSPK-----PIPEHQFGWR-----IID 1527
Db 1520 QCSVDEFCSSLTQTYVDSLKAKENLVLTNLRNFGDLVKEMQLGLEGLVPSLSSSCVPD 1579
Qy 1528 PEDLKALLKVLHLCRGIREKALQKQKHLDVITQACLKNKDVAILLENENENOVTRDIV 1587
Db 1580 SSSLSLGDSSFYRALLEQTQDMSLLSNLEGVUSA--NQCSVDEVFCSLQSENLTRK-- 1635
Qy 1588 ENWSVEEQAMEMDLSVL-----QQVEDLERRVASLQVKGWMCPEPASEREDLVYPEHK 1642
Db 1636 ETPSAPAKGVE-ELESLECVYRQSLKLEEKMES-----QGITMKKEIQELEQLLSERQ 1689

RESULT 42
US-09-949-016-6954
; Sequence 6954, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIORITY FILING DATE: 2000-04-14
; PRIOR FILING DATE: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR FILING DATE: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR FILING DATE: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FASTSEQ FOR Windows Version 4.0
; SEQ ID NO 6954
; LENGTH: 919
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-6954

Query Match 3.1%; Score 310; DB 4; Length 919;
Best Local Similarity 19.6%; Pred. No. 3.5e-13;
Matches 202; Conservative 159; Mismatches 357; Indels 312; Gaps 40;

Qy 62 EGISSSDSD-----LEEDEEEDQSIIESEDDSDSEEAQHSNNQVLLHGISDPKADGQ 118
Db 21 EFLCNMDNKLVLWLEEQEAEAFMR-----EFSKEPELM-----PKTPSQ 62
Qy 119 KATEKAQEKRIHQPLAFESQTHSFOSQOKP--QVLSQQLPFIPOSSQAK-----EESVN 173
Db 63 K--NRKKRRLSY-----VQDENRDPTRRLSRKRSRSSQSSRLRSKSDVE 108
Qy 174 KHTSVIQSTGLVSNVVKPLSLVNQAKETMYKLIIVSPDVLKAGNKNTSEESLLTSELRS 233
Db 109 KLAIVVGENSV--LRRVTRAAAAAATAAAPS-----SPTPESTMLTKKPED 158
Qy 234 KREYKQAFPSQ-----KKQESSKSLKVIKAAALSNPKATSSSPANPKQTLNHNHPPTLN 290
Db 159 NHTQCOLVPVVEIGISERQNAEQHVTLQMLSTEPRLTSLPTPA--SATAPTSGIPTSDE 216
Qy 291 ALLGNHPQNGVIOQVIOEAPLALTTKMO-----SKINENIAAASSTPFSSPV--- 339
Db 217 ESTPKKARILLESITVSSLMATPDQPKGQGVGTGRGASKL--RIAQVSPGPRSDPAPPD 274
Qy 340 -----NLST-SGRRTFG---NOTPVMPSA-SPILHSQ-----GKEKAVSN 374
Db 275 SPWREVRVLAIPILPNFSTPTGSRDTSQSVRHSPIAPSSPQVLAQKYSVLVAKQESVVR 334
```


QY 675 NVGNAEFLDNADAKL---LRKLOAQEIARAQAQIKLRLKLOQBOQARVAK---EAKKQQ 727
Db 1341 ---GERDELQTSCKALVSELELRLRAHVKSVEGENEITTKUNGLEKEILGKSESEVLKSM 1398
QY 728 AIMAEEKRKQEQIKIMQOEKIKRIQOI--RMEKELRAQOIILEAKKKKEE-----778
Db 1399 LENLKEDNNKLEQAEEYSKENQFSLVEVSGSKLVDEIEVLKQALKAABERLEIKDR 1458
QY 779 ---AANAKLLEA-----EKRIKEKEMR-----RQAVL--LKHQ 807
Db 1459 DYFELVQTANTNLVEGKLETPLOADHEEDSDRESEMEIVLGEKLERNOYLLERLQEE 1518
QY 808 ERERRRHHMLKAMEARKKAEERLKOER-RDEKRLNKER-----KLEOR 853
Db 1519 KLELSNKLEILQKEMETSVLKDLOQLKESLLSENILKENIDITLKHHSDDTAQLOKT 1578
QY 854 RLELEMAKELKKPNEDMCADQKPLPELPRIPGLVLSGSTFSDCLMVVQVFLRNFGLGF 913
Db 1579 QOELQAKNLAIASDNCPIQEK-----ETSADC-----1608
QY 914 DVNIDVNLVLQGLNLIGDSMEVQDLLVRLLSAAVCDPELITGYKAKTALGEHLNV 973
Db 1609 ---VHPLEEKILLLTBEHQ-----KTNEQEKLIH- 1635
QY 974 GVNDRDNVSEILQIFWEAHC-----GQTELTESIKTKAFOAHTPAOKASVLAFLINELA 1026
Db 1636 ---ERNELEQAQVELKCEVHELMKSMIESKSLSESLOHEKHDTQOQLLALKQOMQVVT 1690
QY 1027 CSKS-----VVSEID---KNIDVMSNLRD-----KW 1050
Db 1691 QEKKELOQTHELTAEDVHLAKENIELGLNFNEAQOKTKKEQCLLNENKLEQSQHRLQC 1750
QY 1051 VVEGKLRLRIH-AKTKGRKRDGGIDLGEQHPLGTPPTPGRKRRKGGSDYDDDDDD 1109
Db 1751 EIEELMSLKDKESALETKESEQKVINLQEMEMVMEELKNSQRTVIAERDQLODD 1810
QY 1110 ---DSDQDQDEDED---EEDKEQKQKKTIDICEDEBQDQAAVEE---LEKQI----1155
Db 1811 LRESVMSIETQDDLKKAQALQOQKQKQELT-----SQISVLQEKISLENQMLNV 1864
QY 1156 ---EKLSKQO--SOYRRKLPD---ASHSLRSVMFG-----PORYRRYVILPRCG 1197
Db 1865 ATVKETLSERDDLQSKQHLSEIETLSLKEKEFALEQAEDKADAAKTIDITEKIS 1924
QY 1198 GIFVEGMESGGLBEIAKERKLAESVQIKEMFETSGSLNCSNTHCEQKEDLKX 1257
Db 1925 NIEEQLLQQAATNLKETLYERESL-----IQCKEQL-----ALANTEHL--RETLKSK 1968
QY 1258 DNT--NLFLQKPGFSKLSKLEVAKMPPESE-----VMTPKPNAGANGCTLSYQNSGKH 1310
Db 1969 DLALGKMEQERDEANKVIALTE--KMSLSEQINENVTTLKEGEKE--TFYIQRPSKQ 2025
QY 1311 SLGVSQSTATQS-----NVEKAD-----SNLNFNTGSSGPGKFPYSLPNDQLLKTITE 1358
Db 2026 QSSSQMBELRESLTKDOLQLEBAEKISEATNEIKULTAK-----ISSLEE 2071
QY 1359 KNRQWFSLLPRTPCDDTSLTHADNSTASLVTPOQPPSKSPSPPTAPLGSSAQNPVGLNP 1418
Db 2072 EILQVASTILNEAVSERENLRH---SKQLVSELEQLSLTLKSRDHAFQASKREKDEAVNK 2128
QY 1419 FALSLPQVKGVSMMGLQFCGWPFGVTSNI--PTLSVPISLGSLGLSEGNGNSFLTSNV 1477
Db 2129 IASLAEETK-----ILTKEMDEFDRDSKESL-----QEQSHLSEEL 2164
QY 1478 ASSKSESVPQNEKATSAQAAVEAKVDPFSPKPIPEEMQFGWRIIDPEDLKALKV 1537
Db 2165 CTYKTELQMLKQKEDINNKLAEKV-----KEVDLLQ-----2197
QY 1538 LHLRGIRKALQKQIQ-----KHLDYITQACLKNKQVAIITELNENENQVTRDIVENWS 1591
Db 2198 -HLSSLKEQLDQIQWELRNEKRLNVELCEKMDIMEKISVLRMLQNEPQEQBEDDVAERMD 2256

QY 1592 VEOQAMENDLSVLQOVEDLERRRVASASLQVKGWMCPEPASREDLAVYFEHKSFTKLCKEH 1651
Db 2257 ILE-----SRNQEIQELMEKISAV-----YSEQHTLLSSLSSELQKETEAH 2297
QY 1652 DGEFTGDEBSAHALER-----KSNPLDIAVTRLADLERNI-----ERRIEEDIA 1697
Db 2298 KHCMLNIKESLSSTLSRSGSLQTEHVKLNLTQTLNKKFVVYRTAAVKEDHSILIKDYE 2357
QY 1698 PGLRVWRALSEARSAQVALCICOOLOKSTAWESIMK--VYCOICRKGDNEEELLLCDG 1755
Db 2358 KDLAAEQGRHDELRLQO---CLEQHGK--WSQSAEELKFCBI---EFLNELLF-----2405
QY 1756 CDKGCHTYCHRPKITTPDGDWFCPACIAKASGQTL-----KIKKLHVKG 1800
Db 2406 -----KCANIIQSVQDDFSEVOVFLNQVSTLOELEHKKGFQWLEBFGDLHVA 2456
QY 1801 KKTWESKKGKVVUTGTDEDEDSASTSSSLKRGNKDLQKRMENSTINLSKQBSFTSVK 1860
Db 2457 KKLSEGMQ-----QENRRIASTIQLLTKRLKAVVQSKIQREITVYLNQ---FEAKL 2504
QY 1861 KPKRDDSQDLALCSMILTEMETHEDANPFLLPVNLKLVPGYKVKVKKPMDFSTIREKLSS 1920
Db 2505 QEKKEQNKEL-----MRRMEHHGSPASVMEEENARLL-----GILKTVQDES---KKLQS 2551
QY 1921 GOYPNLETFALDVRVFDN 1939
Db 2552 ----RIKMLENELNLVKDD 2566

RESULT 44

US-08-056-200-94
; Sequence 94 Application US/08056200
; Patent No. 5616500
; GENERAL INFORMATION:
; APPLICANT: Steinert, Peter M.
; APPLICANT: Lee, Seung-Chul
; APPLICANT: Kim, In-Gyu
; APPLICANT: Chung, Soo-Il
; APPLICANT: Park, Sang-Chul
; TITLE OF INVENTION: Trichohyalin and Transglutaminase-3 and
; TITLE OF INVENTION: Methods of Using Same
; NUMBER OF SEQUENCES: 117
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 620 Newport Center Drive, Sixteenth Floor
; CITY: Newport Beach
; STATE: CA
; COUNTRY: U.S.A.
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/056,200
; FILING DATE: 30-APR-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Fedrick, Michael F.
; REGISTRATION NUMBER: 36,799
; REFERENCE/DOCKET NUMBER: NIH054.001A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (714) 760-0404
; TELEFAX: (714) 760-9502
; INFORMATION FOR SEQ ID NO: 94:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1898 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-056-200-94

QY 1175 SLRSMFGPDRYRRYRWWILPRCGGIFVEGMESEGL-----BEIAK-EREKLUKKAESV-Q 1227
DB 889 QKRWMLQLEERKRRRHTLYAK--PALQEQLRKEQQLQEBEELQREERERKRRQEQERQ 946
QY 1228 IKEEMFETSGDSLNCNTDHCQKED--LKEKDNTNLPKQKPGSFKLSKLEVAKMPE 1285
DB 947 YREE-----EQLQEBEQLLREERKRRRQERERQYRKDKLQO-----KE 987
QY 1286 SEVMTPKP 1293
DB 988 EQLGEEP 995

RESULT 47

US-07-741-940-7
; Sequence 7, Application US/07741940
; Patent No. 5352775
; GENERAL INFORMATION:
; APPLICANT: ALBERTSEN, HANS
; APPLICANT: ANAND, RAKESH
; APPLICANT: CARLSON, MARY
; APPLICANT: GRODEN, JOANNA
; APPLICANT: HEDGE, PHILIP J.
; APPLICANT: JOSLYN, GEORF
; APPLICANT: KINZLER, KENNETH
; APPLICANT: MARKHAM, ALEXANDER F.
; APPLICANT: NAKAMURA, YUSUKE
; APPLICANT: THLIVERIS, ANDREW
; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
; TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
; NUMBER OF SEQUENCES: 94
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner, Birch, McKie & Beckett
; STREET: 1001 G Street, NW
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001-4598

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/741,940
; FILING DATE: 19920109
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Kagan, Sarah A.
; REGISTRATION NUMBER: 32,141
; REFERENCE/DOCKET NUMBER: 1107.035574
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-508-9100
; TELEFAX: 202-508-9299

; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2842 amino acids
; TYPE: AMINO ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: APC

US-07-741-940-7

Query Match 3.0%; Score 305.5; DB 1; Length 2842;
Best Local Similarity 18.5%; Pred. No. 4.1e-12;
Matches 379; Conservative 256; Mismatches 651; Indels 763; Gaps 92;

QY 14 RKNQEQSKNQP-----LDARVDKIKDKKPKKAMSSSSDSGTSST-----SSEG 63
:: |||::

DB 1060 KQEQQRQRNQSTTYPVVTSTDDDKLXFPQHFQOQECVSPYRSGANGSETNRYGSHG 1119
QY 64 ISSSDDDL-BEDEEEEDQSEEDDDSDSEEAQHKSNNQVLLHGLHSIDPKADGQKATE 122
DB 1120 INQNVQSILCQEDDDYEDDKPTNYSERYSEEGHEEERTNYSIKYN----- 1166
QY 123 KAQEKR-IHQPLPLAFESQTHSFQSQQKQPVLSQOLPFIPOSSQAKESVNHK-TSVIQ 180
DB 1167 --EKKRHVDQPTIDYSLKYAT-DIPSSQK-----SFSFSKSSGSGSKTEHMSSE 1215
QY 181 STGL-VSNVK-----PLSLVNOAKKETMVKLIVP-----S 209
DB 1216 NTSFSSNAKQNLHPSSAQSRSQPKAATCKVSSINQETIQYICVEDTICFSRCS 1275
QY 210 PDVLKAG-----NKNTESESSLLTSELRSKREQYKQAPPSQLKQKQESKSLKVAIALS 263
DB 1276 LSSLSAEDEICGNTTQEADSANTLQIAEIK-----KIGTSAEDPVEVPAVSQ 1327
QY 284 NPKATSS-----SPAHPKQTLNHNHPFLTNALLGNHQHNGVIOQAPLAL 313
DB 1328 HPRTKSSRLQSGSSLSSESARHKAVERFSSGAKSP-----SKSGAQTPKSPPEHYVQETPLMF 1383
QY 314 ---TTTKWQSKINENIAA-----SSTPFSSPV 339
DB 1384 SRCTSVSLDSFESRSIASSVQSEPCSGWVGIIISPSDLPSQPGQTPPSRKTTPPPQ 1443
QY 340 NLSTSGRRTPGNQTP-----VMPASPIILH-----SQGEKAVS 373
DB 1444 TAQTK-REVKNKAPTAEKRESGPKQAANAQVQVLPDADTLHATESTPDGFS 1502
QY 374 NNPNVKTQHHSHPAKSLVQFRGTSDI-----PSSKSDSESDNEDEDEDEDEDD 430
DB 1503 SLSL-----ALSLDEPFIQKDELIMPVQVENDNGNETESEQKESNENQKEA 1552
QY 431 ESDDSQSESDSSESDTEGSEEDDDKQDESDDTSEKTSMLNK-----TTSSVKSPSM 488
DB 1553 EK-TIDSEKLLDDSD-----DDIIEICEIISAMPTKSSRKAKPAQATASKLPPVP 1604
QY 489 SLTGHSTP-----ENLHIAKAPG-SAPAAALCSESQSAFLGTSSS-----TLTS 531
DB 1605 ARKPSQLPVYKLLPSQNLQPKQVSVFTPGDDMPRYCVEG-TPINFSTATSLDLTIES 1663
QY 532 SPH-----SOTSKRR-----RVTDERE-----LRIPLEYGHQRETRIN 565
DB 1664 PNELAAAGEGVRGGAQSGEFKRDITPTEGRSTDEAOGKTSSTVIP-----ELDDNKA 1718
QY 566 FGRLOGEVAYAPCGKKLQYPEVVKYLSRNGIND-ISRDNFSSAKIRVGDIFYEARDG 624
DB 1719 EGDILAEICINSAMPKGSHPKPF-RYKK-----IMDVOQASASSA----- 1758
QY 625 PQEQWCLLKEEDVIPTRAMEGRGRPPNPDPR---QRAREBSRMRKRRKRPVNVGNAB- 680
DB 1759 -----PNKQLDGGKKKPTSPVKPIQNTYRTRVRKADSKNKL-NAER 1802
QY 681 -FLDNADAKLLR-----KLOAQE----- 697
DB 1803 VFSDNKDSKQNLKNSKDFNDKLPNNEDVRGSPAFDPSPHYTPIEGTPYCFSRNDSLS 1862
QY 698 -----TARQAAQIKLLRKLQKQCARVA-----KEAKQQAIAAEKRRKQ 738
DB 1863 SLDFDDDDVDLSREKAEALRKAKE-NKSEAKVTSHTSLTSNQQSANKTQAIKQPINRQ 1921
QY 739 KEQIKIMKQ-----EKIKRIQIRME-----KELRAQOILBAKKKK 775
DB 1922 PKPI-LQKQSTFPQSSKDIIPRGAATDEKLQNFALIENTPVCFSHNSSLSLSDIDQENN 1980
QY 776 KEAANAKLLBAEKRIKEKEMRROQAVLLKHQERRRRQHMLMKAMBAKKAEBKER-- 833
DB 1981 KE---NEPIKETEPDPSQGEPSKQA-----SGVAPKSFHYEDTPV 2018
QY 834 --LKQEKDEKRLNKKERKLQBRLELENAKELKKNEEDMCLADQKPLPRLPIGLVLSG 891
DB 2019 CFSRNSLSLSIDSEDDLLQECISSAMPKK-KKPSR-LKGDNEKHS----- 2064

```
QY 892 STFSCLMVQFLRNFGKVLGFDVNDVPL-----SVLQEGLLNIG 933
Db 2065 -----RNMGGILGEDTLDKDQRPDSEHGLSPDSENFWDKAIQEGANSIV 2111
QY 934 DSMGEVDLLVRLLSAAVC-----DPLITGYKAKTALGEHLNVLGVNRDNNVSEILQI 986
Db 2112 SSLHQ-----AAAACLQRASSDSLSLSKGISLGSPP-----2147
QY 987 PNEAHGQOTELTESLTKAFQAH-----TPAQKASVLAFLINELACSKSVVSEIDKNID 1040
Db 2148 -----HLTPDQEEKPFTSNKGPRLKPGKSTLE-----TKKIESE-----2183
QY 1041 YMSNLRDRKVVVEGKLRLRIIHAKTGKRDTSGGIDLGEHQHPLGTPTGKRKRRKGGD 1100
Db 2184 -----SKGIGKK-----KYKSLITGKVRNSSEIS-GQMQLPQANPSPISGR-----2227
QY 1101 SDYDDDDDDDDQDEDEDEDEKQKG--KTDICEDEDEGQD-----ASVEELEK 1153
Db 2228 -----TWIHIPGVNRSSSTSPVSKKGPPLKTPASKSPSEGQTATTSFPGAKPSVKS 2279
QY 1154 QIEKLSKQSQYRRKLPDASHLSRVMFGPDRYRRYWLPRCGGIFVEGHESGEGLEBI 1213
Db 2280 ELSVVARQTSQ-----NLFYTGSSGPKFVSPPLNDQLLTKLTKRQWFSLLPRTPCDDTS 1376
QY 1214 AKEREKKAESVQIKEMFETSGDLSNCSNTDHCQEKEDLKEKONTNLFQKPGSPS-- 1271
Db 2303 -----SRDS-----TPSRPAQOPLSRP-----IQSPGRNIS 2329
QY 1272 ----KLSKLEAVAKMPPESEVMTPKPNAGANGCTLSYQNSGKHSIGSVQSTATQSNVKA 1327
Db 2330 PGRNGISPPNKLSQLPRTSSPSTASTKSSGSG-KMSYTPSPGRQM--SQNLTKQTGLSKN 2386
QY 1328 DSN-----NLFYTGSSGPKFVSPPLNDQLLTKLTKRQWFSLLPRTPCDDTS 1376
Db 2387 ASSIPRSESASKGLNOMNNGANKKVE-----LSRMSSTKSS 2424
QY 1377 LTHADMSTASLVTPQSPKSPSPPT-----PA-PLGSSAQNPVGL 1416
Db 2425 GSEDRSERPVLVRSTTIKAPSTLRRKLEESASFESLSPSPSPASPTRSQATPV-- 2482
QY 1417 NPFALSLQVKGVSVM-MGLFCGQFPTGVVTSNIPFTLSVPSLGLGSEGNGNSFLTS 1475
Db 2483 ----LSPSLPDMSLSTHSSVQAGW-----RKLPNLS-PT-----IYNDGRPAKRH 2525
QY 1476 NVASKSESP--VPQNEKAT-----SAQPAV 1500
Db 2526 DIARSHSESPRLPTNRSGLTWKREKHSLSPLRVSTWRTGSSSSILSASESEKAKS 2585
QY 1501 EVAKPVDVFPKPIPEEMQF--GWMRIIDPEDLKALLKVLH-----LRGIREKALQK 1551
Db 2586 EDEKRVNLSIGTKQSKENQVSAKGTWRKIKENEFPNTNSTQTVSSSGATNGAESTLIYQ 2645
QY 1552 IOKHLDYITQACLKNKOVAI-IE-----LNENEE 1579
Db 2646 -----MAPAVSKTEDVWVRIEDCPINNPRSGRSPGTNTPPVDVSEKANPNIKDSK 2698
QY 1580 NOVTRDIVENSVBEQ--AMEMDLVLOQVEDLERRVASASLQVK-GWMCPEPASERDL 1636
Db 2699 NQAKQN-VNGSVPMRTVGLNRLNSFTIQVDAPQK----GTEIKPGQNNPVPVSETNES 2753
QY 1637 VYFEHKST 1645
Db 2754 SIVERTPPFS 2762
```

RESULT 48

US-08-289-548A-7

; Sequence 7, Application US/08289548A

; Patent No. 5648212

; GENERAL INFORMATION:

; APPLICANT: ALBERTSEN, HANS

; APPLICANT: ANAND, RAKESH

```
; APPLICANT: CARLSON, MARY
; APPLICANT: GRODEN, JOANNA
; APPLICANT: HEDGE, PHILIP J.
; APPLICANT: JOSLYN, GEOFF
; APPLICANT: KINZLER, KENNETH
; APPLICANT: MARKHAM, ALEXANDER F.
; APPLICANT: NAKAMURA, YUSUKE
; APPLICANT: THLIVERIS, ANDREW
; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
; TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
; NUMBER OF SEQUENCES: 102
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner & Allegretti, LTD
; STREET: 1001 G Street, NW
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001-4598
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/289,548A
; FILING DATE: 12-AUG-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Kagan, Sarah A.
; REGISTRATION NUMBER: 32,141
; REFERENCE/DOCKET NUMBER: 1107.46943
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-508-9100
; TELEFAX: 202-508-9299
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2842 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: APC
; US-08-289-548A-7
```

Query Match 3.0%; Score 305.5; DB 1; Length 2842;

Best Local Similarity 18.5%; Pred. No. 4.1e-12;

Matches 379; Conservative 256; Mismatches 651; Indels 763; Gaps 92;

```
QY 14 RKNQEQSKNOP-----LDARVDKIKDKPRKAMESSNSDSGTSSTDT-----SSEG 63
Db 1060 KQSEQRQSRNQSTTPYVYVTESTDDKHLKQPHQEQECVSPYRGRGANGSETNRVGSNHG 1119
QY 64 ISSSDSDDL-PEDEEEDQSTEESEDDSDSESAQKSNQVLLHGISDPKADQKATE 122
Db 1120 INQVNSQSCQEDYEDDKPTNYSERYSEEEQHHEERPTNYSIKYN-----1166
QY 123 KAQEKRI-HQPLPLAFESQTHSFQSQKQPOVLSQLPFIFQSSQAKEESVNHK--TSVIQ 180
Db 1167 --EKRHVDPQIDYSLKAT-DIPSSQKQ-----SFSPSKSSGSSQSKTEHSSSSE 1215
QY 181 STGL-VSNVK-----PLSLVNQAKETMYMKLIIV-----S 209
Db 1216 NTSTPSSNAKTRQNLHPFSSAQSRSQGPQKAATCVSSINQETIQTCYVEDTPICFSRCSS 1275
QY 210 PDVLKAG-----NKNTSESSLLTSELRSREYKQAFPSQLKQESSKSLKVVIAALS 263
Db 1276 LSSLSSADEIGCQCTQEQADSANTLQAEIKE-----KIGTRSAEDPVSEVPVSVQ 1327
QY 264 NPKATSS-----SPAHPKOTLENNHPNPFITNALLGNHQPNGVIQSVIOBAPLAL 313
```

Db 1328 HPRTKSSRLQSSLSSESARHKAVERFSSCAKSP-----SKSGAQTPKSPPEHVVOETPLMF 1383
Qy 314 -----TTKTRMOSKINENIAAA-----SSTPFSSPV 339
Db 1384 SRCTSVSSLDSPESRSIASSVQSEPCGMVGIISPSDLDPGQTMPPSRKSTPPPPQ 1443
Qy 340 NLSTSGRRTPGNQTP-----VMPASPIIH-----SQGEKAVS 373
Db 1444 TAQTK-REVFGKNAPTAEKRESGPKQAAVNAAVQRVQLPDLADTLHLHFATESTPDGFSCS 1502
Qy 374 NNVPVKTQHHGHPAKSLVEQFRGTDSDI---PSSKSDSDSNEDEDEDEDEDDDD 430
Db 1503 SLSL-----ALSLDEPIQKDVLRIMPVQENDNGNETSEBQPKESNENQKEA 1552
Qy 431 ESDSDSDSDSSESDTEGSEEDDDKQDQDESDDTEGEKTSMLNK--TTSSVKSPSM 488
Db 1553 EK-TIDSEKDLDDSD-----DDIEILEECIISAMPTKSRKAKPAQATASKLPPPV 1604
Qy 489 SLTGHSTP-----RNLHTAKAPG-SAPAALCESOSPAFLGTSSS-----TLTS 531
Db 1605 AKKPSQLPVYKLLPSQNLQPKQVHSFTPGDDMPRYCVEG-TPINFSTATSLSDLTIES 1663
Qy 532 SPH-----SGTSKR-----RVTDERE-----LRIPLEYGMORETRIRN 565
Db 1664 PNELAAAGEVGRGAQSQGEFEKRDPTPEGRSTDEAQGKTSSTVIP-----ELDNKAE 1718
Qy 566 FGRIGQGEVAYVAPCGKLRQYPIVKYLSRNGIMD-ISRDNFSPSAKIRVGDIFYEARDG 624
Db 1719 EGDILAECINSAMPKSHKPP-RVKK-----IMDQVQASASSA----- 1758
Qy 625 POEMOWCLLEEDVTPRIAMGRGRPNPNDR-----QRARESRMRRRGRPPNVGNAB- 680
Db 1759 -----PNKNQLDGKKKPTSPVKPIQNTYRTRVRKNADSKNKL-NAER 1802
Qy 681 -FLDNADAKLLR-----KLQAE----- 697
Db 1803 VPSDNKSKQNLKNNSKDFNKLNNEDRVGRGSAFDSPHYPIEGTPYCFSRNDSLS 1862
Qy 698 -----IARQAQIKLRLKQEQARVA-----KEAKQQAIIAABEKRAQ 738
Db 1863 SLDFDDDDVDLSREKAEALRKAKE-NKESEAKVTSHTLTSNOQSANKTQAIKQPINRQ 1921
Qy 739 KEQIKIMQO-----EKIKRIQIRM-----KEIRAQOILEAKKKK 775
Db 1922 PKPI-LOKQSTFPQSSKOIPDRGAATDEKLQNFATIENTPVCFSHNSLSLSLSDIQENNN 1980
Qy 776 KEAANAKLLEAKRIKEKEMERQOAVLLKHOERERRRQHMMLKAMEARKKAEKER-- 833
Db 1981 KE---NEPIKETPPDSQGEPSKQA-----SGYAPKSFHVEDTPV 2018
Qy 834 --LKOEKREKLNKERKLEQRRLLEMAKELKPNEDMCLADOKPLPRLPIGLVLGG 891
Db 2019 CFSRNSLSLSLSDSEDDLQECISSAMPKK-KKPSR-LKGDNEKHP----- 2064
Qy 892 STFSCLMVQPLRNFQKVLGFDVNDVNL-----SVLQEGLLNIG 933
Db 2065 -----RNMGGILGSDTLDLKDIQRPDSEHGLSPDSENFOWKAIQEGANSIV 2111
Qy 934 DSMGEVQDLLVPLLSAANC-----DPGLITGYAKTALGEHLNNGVNRDNNVSEILQI 986
Db 2112 SSLHOA-----AAAACLSRQASDSDSLSLSKSGISLSPF----- 2147
Qy 987 FMEAHCGOTELTESLTKTAFOAH-----TPAQKASVLAFLINELACSKSVVSEIDKNID 1040
Db 2148 -----HLTPDQEEKPFTSNKGRPLIKPEKSTLE-----TKKIESE----- 2183
Qy 1041 YMSNLRRDKVWVEKRLKRLRIIHAQTKGRDTSGGIDLGEEQHPLGTPTPGRKRRKGGD 1100
Db 2184 -----SKGIKGGK---KVYKSLITGKVRNSSEIS-GQMKQPLQANNMPSISRGK---- 2227
Qy 1101 SDYDDDDDDSDQDE 1153
Db 2228 -----TMTHIPGVRNSSSSTSPVSKKGPPLKTPASKSPSEGTATTSRGAKPSPVKS 2279

Qy 1154 QIEKLSKQOQSVRRKFLDASHSLRSVMFGPDYRRRYWILPCGGIFVEGMSGEGLEBI 1213
Db 2280 ELSPVARTSQ-----IGSSSKAPSRSG----- 2302
Qy 1214 AKEREKLLKAESVQIKEEMFETSGDLSLNCNTDHCQEOKEDLKEKONTNLFLOKPGSFS-- 1271
Db 2303 -----SRDS-----TPSRPAOQLSRP-----IOSPGRNSIS 2329
Qy 1272 ----KLKLLLEVAKMPPESEVMTPKPNAGANGCTILSYQNSGKHSLGSGVSTATQSNVEKA 1327
Db 2330 PCBNGISPNKLSQLPRTSSPSTASTKSGSG-KMSYTPSGROM--SQNLTKQICLSN 2386
Qy 1328 DSN-----NLFMTGSGGPKFYSPLPNDOLLKTLTEKNQWFSLLPRTPCDDTS 1376
Db 2387 ASSIPRESASAKGLNMNGNKGANKKVE-----LSRMSSTKSS 2424
Qy 1377 LTHADMSTASLVTPQSPKSPSP-----PA-PLGSSAQNIVCL 1416
Db 2425 GSEDRSERPVLVROSTFIKEAPSTLRRKLEESASFESLSPSPSPASPTRSOAOTPV-- 2482
Qy 1417 NPFALSPQVKGVSMM-MGLOFCGWPTGVVTSNIFTLSVPSLGSGLGSLSEGNNGSFLT 1475
Db 2483 ----LSPSLPDMSLSTHSSVQAGW-----RKLPNLS-PT-----LEYNDGRPAKH 2525
Qy 1476 NVASSKSESP--VPQNEKAT-----SAQPAV 1500
Db 2526 DIARSHSESPRLPINRSGTWKREHSHSSLPVSTWRRTGSSSSSILSASSESEKAKS 2585
Qy 1501 EVAKVDVFPKPIPEEMQF---GHWRIIDEDDLKALLKVLH-----LRGREKALQK 1551
Db 2586 EDEKHNISIGTKQSKENQVSAKGTWRKIKENEFPTNSTQTVSSGATNGAESKTLIYQ 2645
Qy 1552 IQKHLDTYOACLKNDKVAI-IE-----LNENEE 1579
Db 2646 -----NAPAVSKTEDVWRIEDCPINNPSGRSPTGNTPPVIVSVSEKAMPNIXKD 2698
Qy 1580 NQVTRDIYENMSVEBQ--AMEMDLSVLOQVEDLERRVASASIQVK-GMMCPEPASEREDL 1636
Db 2699 NOAKQN-VNGSVPMRTVGLNRLNSFIQVDAPDK---GTEIKPGQNNPVPVSETNES 2753
Qy 1637 VYFEHKSFT 1645
Db 2754 SIVERTPFS 2762

RESULT 49
US-08-452-654-7
; Sequence 7, Application US/08452654
; Patent No. 5691454
; GENERAL INFORMATION:
; APPLICANT: ALBERTSEN, HANS
; APPLICANT: ANAND, RAKESH
; APPLICANT: CARLSON, MARY
; APPLICANT: GRODEN, JOANNA
; APPLICANT: HEDGE, PHILIP J.
; APPLICANT: JOSLYN, GEOFF
; APPLICANT: KINZLER, KENNETH
; APPLICANT: MARKHAM, ALEXANDER F.
; APPLICANT: NAKAMURA, YUSUKE
; APPLICANT: THLIVERIS, ANDREW
; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
; NUMBER OF SEQUENCES: 94
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner, Birch, McKie & Beckett
; STREET: 1001 G Street, NW
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001-4598
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/452,654

FILING DATE: 25-MAY-1995

CLASSIFICATION: 536

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/741,940

FILING DATE: 08-AUG-1991

ATTORNEY/AGENT INFORMATION:

NAME: Kagan, Sarah A.

REGISTRATION NUMBER: 32,141

REFERENCE/DOCKET NUMBER: 1107.035574

TELECOMMUNICATION INFORMATION:

TELEPHONE: 202-508-9100

TELEFAX: 202-508-9299

INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:

LENGTH: 2842 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

ORIGINAL SOURCE:

ORGANISM: Homo sapiens

IMMEDIATE SOURCE:

CLONE: APC

US-08-452-654-7

Query Match 3.0%; Score 305.5; DB 1; Length 2842;

Best Local Similarity 18.5%; Pred. No. 4.1e-12;

Matches 379; Conservative 256; Mismatches 651; Indels 763; Gaps 92;

QY 14 RKCNEQSKNP-----LDARVDKIDKPKRKAMSSSSDSGTSST-----SSEG 63

DB 1060 KQSEQRQSRNOSTTTPVYTESTDDKHLKFQPHFGQCECVSPYRSRGANGSETNRVGSNHG 1119

QY 64 ISSSDSDLL-REDEEEQSTIESEDDSDSESEAHQKSNQVLLHGISDPKADQXKATE 122

DB 1120 INQNVQSQCQEDDYDDKPTNYSRYSEEEQHEEERTNYSKYN-----1166

QY 123 KAQEK-THQPLAFESGTHSFQSQKQPVLSQQLPFIQSSQAKESVNKH-TSVIQ 180

DB 1167 --EERKHVDQIDYSLKAT-DIPSSQK-----SFSFKSSGSSQSKTEHSSSSSE 1215

QY 181 STGL-VSNVK-----PLSLVNAKQKTYMKLIIVP-----S 209

DB 1216 NTSTPSSNAKRNQLHPSSAQSRSGQPQKAATCKVSSINQETIQYCYVEDTPICFSRCSS 1275

QY 210 PDVLKAG-----NKNVTESSLLTSELRSKREQYKQAPPQOLKQESKSLKKVIALS 263

DB 1276 LSSLSAEDIEGCTQTQEADSANTLQIAETKE-----KIGTRSAEDPVSEVPAVSQ 1327

QY 264 NPKATSS-----SPAHPKQTLNHNHPFLTNALLGNHOPNGVIOQSVIOEAPLAL 313

DB 1328 HPRTKSSRLQSSLSSESARHKAVEFFSGAKSP-----SKSGAQTPKSPPEHYVQETPLMF 1383

QY 314 ---TTKTKVQSKINENIAAA-----SSTPFSSPV 339

DB 1384 SRCTSVSSLDSPESRSIASSVQSEPCSGMVGSIISPSLDPSPGQTMPPSRKTPPPPPQ 1443

QY 340 NLSTSGRTPGNQT-----VMPASPLH-----SQGKEKAVS 373

DB 1444 TAQTK-REVPKRNKAPTAEKRESGPKQAANAARVQVQLPADTLHLHFATESTPDGFS 1502

QY 374 NNVPVKTQHHSHPAKSLVEQFRGTDSDI---PSSKQSDSENEDEEEDDEEDDD 430

DB 1503 SLSL-----ALSLDEFIQKQVELRIMPVQENDNGNETESEQPKESNENQEK 1552

QY 431 ESDSQSSDSNESDTEGSEEEEDDDKQDESDSDTEGKYSMKLNK--TTSSVKSFSM 488

DB 1553 EK-TIDSEKOLLDDSD-----DDDIIEECIIISAMPTKSSRKAKKPAQTASKLPPPV 1604

QY 489 SLTGHSTP-----RNLIHAKAPG-SAPAAALCSESQSAPFLCTSSS---TLTSS 531

DB 1605 ARKESQLPVYKLLPSQNRLOPKHVSFTPGDDMRVYCVEG-TPINFSTATSLDLTTES 1663

QY 532 SPH-----SGTSKRR-----RVTDERE-----LRIPLEYGWQRETRIRN 565

DB 1664 PPNELAAAGEVGRGAQSGEPEKRTDIPTEGRSTDEAQGGKTSSTVIP-----ELDDNKA 1718

QY 566 FGRLOGEVAYAPCGKKLROYPEVIKYLRSNGIMD-ISRDNFSAKIRVGDIFYEARDG 624

DB 1719 EGDILAEICINSAMPKGSHPF-RVYK-----IMDVQQAASASSA-----1758

QY 625 PQEMQWCLLKEEDVIPRIAMEGRGRPPNDR---QRAEESMRKRKGRPPNVGNAE- 680

DB 1759 -----PNKNQLDGKKKPTSPVKPIQNTYRTRVRKNADSKNNL-NAER 1802

QY 681 -FLDNADAKLLR-----KLQAE-----697

DB 1803 VFSNDKSKQNLKXNSKDFNDKLPNNEDVRGSPAFDPSPHYTPTECTPYCFSRNDSLS 1862

QY 698 -----IARQAAQIKLLRKLQOEQARVA-----KEAKKQQAIAAEEKQK 738

DB 1863 SLDFDDDDVDLSREKAEILKAKE-NKESEAKVTSHTELTSNQOSANKTQAIAKQPINRG 1921

QY 739 KEQIKIMKQ-----EKIKRIQIIRME-----KELRAQOILLEAKKKK 775

DB 1922 PKPI-LOKQSTFPOSSKDIPDRGAATDEKLNFAIENTPVCFSHNSLSUSDIDQENN 1980

QY 776 KEEAANAKLLEAEKRIKEKEMRQOAVLLKHQERRRRQHMLMKAMARKAKEKER-- 833

DB 1981 KE---NEPIKETEPDPSQGEPSKQA-----SGVAPKSFHVEDTPV 2018

QY 834 --LQEKREDEKRLNKRLEQRRLLEMAKELKPNEDMCLADOKPLPELPRIPELVLSG 891

DB 2019 CFSRNSLSLSIDSEDDLQECISSAMPKK-KKPSR-LKGDNEKHSF-----2064

QY 892 STPFDCLMVVQFLNFGKVLGFDVNDVNL-----SVLQEGCLNIG 933

DB 2065 -----RNMGGILGEDLTLDLQIQRDPSEHGLSPDSENFDMWKAIOEGANSIV 2111

QY 934 DSMGEVDLLVRLLSAIVC-----DPGLITGYKAKTALGEHLNVLNVDNRNVSEILQI 986

DB 2112 SSLHQA-----AAACLRSQASDSDSLSLKSGISLGSPPF-----2147

QY 987 FMEAHGQTELTESLTKYAFQAH-----TPAQKASVLAFILNELACSKSVVSEIDKNID 1040

DB 2148 -----HLTPDQEEKPFTSNKGPRIKPGKSTLE-----TKKIESE-----2183

QY 1041 YMSNLRDKWVVEGKRLRIIHAKTCKRDTSGDIDLGERQHPLGTPTRGRKRKRGD 1100

DB 2184 -----SKHGKGK---KVYKSLITGKVRNSSEIS-GOMKQPLQANMPSISRGR---2227

QY 1101 SDYDDDDDDDDQDEDEDEKQKQK-KTDICEDDEGDQA-----ASVEELEK 1153

DB 2228 -----TMHIHGVNRNSSTSPVSKGPGPLKTPASKSPSEGTATTSRGAKPVSYS 2279

QY 1154 QIEKLSKQOQSYRRKLFDASHLSRVMFGPDRYRRYWLPRCGGIFVEGMESGGLBEI 1213

DB 2280 ELSPVARQTSQ-----IGGSSKAPSRSG-----2302

QY 1214 AKEREKLKKAESVQI KEEMFETSGDSLNCNSNTDHCQEKDLKXDNNTNLFQKQGSFS-- 1271

DB 2303 -----SRDS-----TPSRPAQQPLSRP-----IQSPGRNIEIS 2329

QY 1272 ---KLSKLLVAKMPPESEVWTPKPNAGANGCTLSYNSGKHSIGSVQSTATQSNVEKA 1327

DB 2330 PGRNGISPPNKLSQLPRTSSPSTASTKSSGSG-KMSYTSRQOM--SQONLTKQTGLSKN 2386

QY 1328 DSN-----NLFNTGSSGPKFYPSLPNDQLLTKLTKNRQWFSLLPRTPCDDTS 1376

DB 2387 ASSIPRSASAKGLNMNNGNANKKVE-----LSRMSSTKSS 2424


```
Qy 681 -FLDNADAKLLR-----KLOAQE----- 697
| | | | |
Db 1803 VFSNKKSKQNLKNNKDFNDKLPNEDRVGSPAFDPHHYPTPIETGYCFPSRNDLS 1862
| | | | |
Qy 698 -----TARQAAQIKLRLKLOKQCARVA-----KEAKQQAIAAAEERKQ 738
| | | | |
Db 1863 SLDFDDDDVLSREKAEKAKE-NKESEAKVTSHTELTSNQQSANKTAIAKQPINRGQ 1921
| | | | |
Qy 739 KEQIKIMKQ-----EKIKRIQOIRME-----KELRQAQILAEAKKK 775
| | | | |
Db 1922 PKPI-LQKQSTFPOSSKDIPRGAATDEKLNQFAIENTVCFSHNSLSLSDIDQENNN 1980
| | | | |
Qy 776 KEEAANAALLAEKHIKEKEMRROQAVLLKQERERRQHMLMKAMEARKKAEKER-- 833
| | | | |
Db 1981 KE---NEPIKETPPDPSQGEPSKPA-----SGYAPKSFHVHEDTPV 2018
| | | | |
Qy 834 --LKQEKDEKRLKLEQRRLLEMAKELKKENEDMCLADOKPLPELPRIPLGLVLSG 891
| | | | |
Db 2019 CFSRNSLSLSIDSEDDLQECISSAMPKK-KKPSR-LKGDNEKHSP----- 2064
| | | | |
Qy 892 STFSCLMVQFLRNFQKVLGFDVNDVNL-----SVLQEGLLNIG 933
| | | | |
Db 2065 -----RNMGGILGEDLTLDLQIDQRPDSEHGLSPDSENFDFWKAIOEGANSIV 2111
| | | | |
Qy 934 DSMGEVQDLLVRLLSAAVC-----DPGLITGVKAKTALGEHLLNVGVNRDNYSEILQI 986
| | | | |
Db 2112 SSLHQA-----AAAACLQRQASSDSLSLSKGISLSPF----- 2147
| | | | |
Qy 987 FMEAHCGOTELTESLTKTAKFAH-----TPAQASVLAFLINELACSVSVSEIDKNID 1040
| | | | |
Db 2148 -----HLTPDQEEKPFTSNKGPRLTKPGEKSTLE-----TKKIESE----- 2183
| | | | |
Qy 1041 YMSNLRDKWVVEGLKRLRIIHAKTGKRTSGGIDLGEBOHPLGTTPGKRRRKGGD 1100
| | | | |
Db 2184 -----SKGIGKGR-----KVYKSLITGKVRNSNIS-GOMKQPLQANMPSISGR----- 2227
| | | | |
Qy 1101 SDYDDDDDDDDQDEDEDEEDKQK--KTDICEDDEGDQA-----ASVEELEX 1153
| | | | |
Db 2228 -----TMHIPGVRNSSSTSPVSKKGPLKTPASKPSSEGQTATTSPRGAKPSVKS 2279
| | | | |
Qy 1154 QIEKLSQOQSVRRKLPDASHLSRSMFGPDRYRRYILPRCGGIFVEGMESGEGLEEI 1213
| | | | |
Db 2280 ELSPVARTSQ-----SRDS-----TPSRPAQQPLSRP-----IQSPGRNSIS 2329
| | | | |
Qy 1214 AKEREKLLKASVQIKEMFETSGDSLNCSTNDHCEQKEDLKEKDNNTLFLQKPGSFS-- 1271
| | | | |
Db 2303 -----KLSKLLLEVAKMPPSEVMTKPNAGANGCTLSYQNSGKHSLSGVSQSTATQSNVEKA 1327
| | | | |
Qy 1272 ---KLSKLLLEVAKMPPSEVMTKPNAGANGCTLSYQNSGKHSLSGVSQSTATQSNVEKA 1327
| | | | |
Db 2330 PGRNGISPPNKLUSQLPRTSSPSTASTKSGSG-RMSYTSRQOM--SQONLTQKQGLSKN 2386
| | | | |
Qy 1328 DSN-----NLFTNGSSGPGKFSPLPNDQLLTKLTKNKNQWFPLLPRTPCDDTS 1376
| | | | |
Db 2387 ASSIPRSESASKGLNQMNNGANKKVE-----LSRMSSTKSS 2424
| | | | |
Qy 1377 LTHADMSTASLVTPOSQPPSPSPST-----PA-PLGSSAQNPVGL 1416
| | | | |
Db 2425 GSEDRSERPVLVRQSTFIKEAPSPTLRRKLEESASFESLSPSSRPASPSTRSQATPV-- 2482
| | | | |
Qy 1417 NPFALSPLOVKGYSM-MGLQFCWPTGVVTSNIPFTLSVPGLSGLSGLSEGNNSFLTS 1475
| | | | |
Db 2483 ----LSPSLPDMJSJTHSSVQAGW-----RKLPPLNLS-PT-----IYNDGRGPAKRH 2525
| | | | |
Qy 1476 NVASSKSESP--VPONEKAT-----SAQPAAV 1500
| | | | |
Db 2526 DIARSHSESPRLPLNRSCTWKREHSKHSSSLPRVSTWRTGSSSSILSASESEKAKS 2585
| | | | |
Qy 1501 EVAKPVPDFPSPKPIPEEQF---GWWRIIDPEDLKALLKVLH-----LRGIREKALQK 1551
| | | | |
Db 2586 EDEKHNISIGTKQSKENQVSAKQGTWRKI KENEFSPNTNSTQTVSSGATNGAESKTLIYQ 2645
| | | | |
```

```
Qy 1552 IQKHLDYITQACLKNKDVAI-IE-----LNENEE 1579
| | | | |
Db 2646 -----MAPAVSKTEDVWVRIEDCPINNPRSGRSPGTNTPPVIDSVSEKANPNIKOSKD 2698
| | | | |
Qy 1580 NOVTRDIVENKSVBEQ--AMENDLSVLQOVEDLERRVASASLQVK-GWMCPEPASERDDL 1636
| | | | |
Db 2699 NOAKQN-VGNGSVPMRTVGLNRLNSFTQVDAPQK-----GTEIKPGQNNPVPVSETNES 2753
| | | | |
Qy 1637 VYFEHKSFT 1645
| | | | |
Db 2754 SIVERTPFS 2762
| | | | |
```

Search completed: March 9, 2005, 14:59:00
Job time : 122.697 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 9, 2005, 13:18:56 ; Search time 186.814 Seconds
(without alignments)
5405.490 Million cell updates/sec

Title: US-10-702-148-21

Perfect score: 10138

Sequence: 1 MGQTKSTSGGNKCNQEQ.....AGHNMRYFKKWTDTFKVS 1972

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 65 summaries

Database :

UniProt_03.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	10097	99.6	1972	1 BA2B_HUMAN	Q9uif8 homo sapien
2	8138.5	80.3	2130	1 BA2B_CHICK	Q9del3 gallus gall
3	5274	52.0	1369	2 Q9NZF8	Q9NZF8 mus musculu
4	4153	41.0	795	2 Q9NC87	Q9NC87 homo sapien
5	3093	30.5	861	2 Q9C0K4	Q9C0K4 mus musculu
6	2515.5	24.8	1905	2 Q9BDI8	Q9BDI8 homo sapien
7	2513.5	24.8	1878	1 BA2A_HUMAN	Q9uif9 homo sapien
8	2437	24.0	1850	1 BA2A_MOUSE	Q9iye5 mus musculu
9	2386.5	23.5	1698	2 Q9Y194	Q9Y194 xenopus lae
10	1543	15.2	2060	2 Q7Q359	Q7Q359 anopheles g
11	1466.5	14.5	3080	2 Q9V602	Q9V602 drosophila
12	1452	14.3	3109	2 Q9BMQ0	Q9bmq0 drosophila
13	1331	13.1	886	2 Q8OU42	Q8OU42 mus musculu
14	1169	11.5	441	2 Q6MZK7	Q6mzk7 homo sapien
15	1090	10.8	1376	2 Q23590	Q23590 caenorhabdi
16	721.5	7.1	1556	1 BA1A_HUMAN	Q9nrl2 h bromodoma
17	639	6.3	131	2 Q8CFE4	Q8cfp4 mus musculu
18	562	5.5	769	2 Q8CGH2	Q8cgh2 mus musculu
19	558	5.5	683	2 Q96013	Q96013 drosophila
20	536	5.3	1483	1 BA1B_HUMAN	Q9uigo homo sapien
21	527	5.2	1079	2 Q8UVF4	Q8uvr4 xenopus lae
22	517	5.1	1357	2 Q7PRP9	Q7prp9 anopheles g
23	513	5.1	1479	1 BA1B_MOUSE	Q9z277 mus musculu
24	511.5	5.0	709	2 Q8OV18	Q8ov18 mus musculu
25	481	4.7	1476	2 Q9NG24	Q9ng24 drosophila
26	481	4.7	1476	2 Q9V974	Q9v974 drosophila
27	479	4.7	1476	2 Q9Y0W1	Q9y0w1 drosophila
28	474.5	4.7	892	2 Q7P199	Q7p1d9 mus musculu
29	458.5	4.5	1508	2 Q75JP5	Q75jps dictyosteli
30	449.5	4.4	1427	2 Q9N519	Q9n519 caenorhabdi
31	419	4.1	657	2 Q8CAU9	Q8cau9 mus musculu

32	417.5	4.1	718	2	Q8C8D1	Q8c8d1 mus musculu
33	409.5	4.0	2095	2	Q7RN75	Q7rn75 plasmodium
34	402	4.0	7210	2	Q9V7G8	Q9v7g8 drosophila
35	402	4.0	9270	2	Q8MLD9	Q8mld9 drosophila
36	397	3.9	18519	2	Q8ISF6	Q8isf6 caenorhabdi
37	397	3.9	18534	2	Q8ISF7	Q8isf7 caenorhabdi
38	393.5	3.9	181	2	Q8BRF6	Q8brp6 mus musculu
39	391.5	3.9	2081	2	Q9LH98	Q9lh98 arabidopsis
40	390.5	3.9	2759	2	Q9VID9	Q9vid9 drosophila
41	390	3.8	2781	2	Q9UIG2	Q9uig2 homo sapien
42	384.5	3.8	1268	2	Q6CGN4	Q6cgn4 varrovia li
43	381	3.8	2760	2	Q8ISY2	Q8isy2 plasmodium
44	378.5	3.7	3259	1	GOBL_HUMAN	Q14789 homo sapien
45	377.5	3.7	627	1	BA1A_XENLA	Q8uvr5 xenopus lae
46	376	3.7	5412	1	FUTS_DROME	Q9w596 drosophila
47	373	3.7	1811	2	Q7R6D1	Q7r6d1 giardia lam
48	369	3.6	460	2	Q8MKZ9	Q8mkz9 drosophila
49	365.5	3.6	3111	2	Q9VH10	Q9vhl0 drosophila
50	365	3.6	1661	2	Q06166	Q06166 plasmodium
51	364.5	3.6	2764	2	Q7Z7D6	Q7z7d6 homo sapien
52	362.5	3.6	1418	2	Q86HT4	Q86ht4 dictyosteli
53	362.5	3.6	2910	2	Q9FND5	Q9fnd5 arabidopsis
54	359.5	3.5	1202	2	Q7QBK2	Q7qbk2 anopheles g
55	359.5	3.5	3257	2	Q9V736	Q9v736 drosophila
56	358	3.5	1909	2	Q25893	Q25893 plasmodium
57	357	3.5	1564	2	Q7Z7J3	Q7z7j3 homo sapien
58	355.5	3.5	10578	2	Q8ISF5	Q8isf5 caenorhabdi
59	354	3.5	162	2	Q6PE75	Q6pe75 mus musculu
60	353.5	3.5	2479	2	Q7SDD4	Q7sdd4 neurospora
61	353.5	3.5	16223	2	Q8IR22	Q8ir22 drosophila
62	353	3.5	249	2	Q9ST46	Q9st46 drosophila
63	352	3.5	2392	2	Q810K1	Q810k1 caenorhabdi
64	352	3.5	2407	2	Q7YSH5	Q7ysh5 caenorhabdi
65	351.5	3.5	3404	2	Q7RJ33	Q7rjb3 plasmodium

ALIGNMENTS

RESULT 1

BA2B_HUMAN

ID BA2B_HUMAN STANDARD; PRT; 1972 AA.

AC Q9UIF8; Q96EAL; Q96SQ8; Q9P252; Q9Y4N8;

DT 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 05-JUL-2004 (Rel. 44, Last annotation update)

DE Bromodomain adjacent to zinc finger domain 2B (hwalp4).

GN Name=BAZ2B; Synonyms=KIAA1476;

OS Homo sapiens (Human)

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Carnivora; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A. (ISOFORM 1).

RC TISSUE=Testis;

RA Jones M.H., Hamana N., Nezu J., Shimane M.;

RT "A novel family of bromodomain genes.";

RL Genomics 63:40-45(2000).

RN [2]

RP SEQUENCE FROM N.A. (ISOFORM 3).

RC Ohara O., Nagase T., Kikuno R.;

RA Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.

RN [3]

RP SEQUENCE OF 753-1972 FROM N.A.

RC MEDLINE=20277482; PubMed=10819331;

RA Nagase T., Kikuno R., Ishikawa K.-I., Hirose M., Ohara O.;

RT "Prediction of the coding sequences of unidentified human genes. XVII.

RT The complete sequences of 100 new cDNA clones from brain which code

for large proteins in vitro."

RL DNA Res. 7:143-150(2000).

RN [4]

RP SEQUENCE OF 1-776 FROM N.A. (ISOFORM 1).

RC TISSUE=Melanoma;
 RL Ansgorge W., Wirkner U., Mewes H.-W., Weill B., Wiemann S.;
 RA Submitted (JUL-2002) to the ENBL/GenBank/DBJ databases.
 RN [5].
 RP SEQUENCE OF 1-799 FROM N.A. (ISOFORM 2).
 RC TISSUE=Skeletal muscle;
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins P.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heide F.,
 RA Daplatenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Staptchenko M., Soares M.B., Bonaldi M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,
 RA Boak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Huiyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahy J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalls D.B.,
 RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 RN [6].
 RP SEQUENCE OF 427-1215 FROM N.A. (ISOFORM 3).
 RX PubMed=14702039; DOI=10.1038/ng1285;
 RA Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,
 RA Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,
 RA Sekine M., Ohayashi M., Nishi T., Shibahara T., Tanaka T., Ishii S.,
 RA Yamamoto J.-I., Saito K., Kawai Y., Isono Y., Nakamura Y.,
 RA Nagahari A., Sudo H., Hosioki T., Kaku Y., Kodaira H., Kondo H.,
 RA Sugawara M., Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E.,
 RA Omura Y., Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa M.,
 RA Yamazaki M., Ninomiya K., Ishibashi T., Yamashita H., Murakawa K.,
 RA Fujimori K., Tanai H., Kimata M., Watanabe M., Hiraoa S., Chiba Y.,
 RA Ishida S., Ono Y., Takiguchi S., Watanabe S., Yosida M., Houta T.,
 RA Kusano J., Kanehori K., Takahashi-Fujii A., Hara H., Tanase T.-O.,
 RA Nomura Y., Togai Y., Komai F., Hara R., Takeuchi K., Arita M.,
 RA Inose N., Musashino K., Yuuki H., Oshima A., Sasaki K., Aota S.,
 RA Yoshikawa Y., Matsunawa H., Ichihara T., Shiohata N., Sano S.,
 RA Moriya S., Momiyama H., Satoh N., Takami S., Terashima Y., Suzuki O.,
 RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,
 RA Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,
 RA Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,
 RA Fujimori Y., Koniyama M., Taehiro H., Tanigami A., Fujiwara T.,
 RA Ono T., Yamada K., Fujii Y., Ozaki K., Hiro M., Ohmori Y.,
 RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
 RA Okitani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,
 RA Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,
 RA Togashi T., Oyama M., Hata H., Watanabe M., Komatsu T.,
 RA Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,
 RA Okumura K., Nagase T., Nomura N., Kikuchi H., Masuho Y., Yamashita R.,
 RA Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.;
 RT "Complete sequencing and characterization of 21,243 full-length human
 RT cDNAs.";
 RL Nat. Genet. 36:40-45 (2004).
 RN [7].
 RP SEQUENCE OF 1524-1972 FROM N.A.
 RC TISSUE=Testis;
 RA Poustka A., Klein M., Mewes H.-W., Gassenhuber J., Wiemann S.;
 RA Submitted (JUN-1999) to the ENBL/GenBank/DBJ databases.
 CC -1- FUNCTION: May play a role in transcriptional regulation
 CC interacting with ISWI.
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event-Alternative splicing; Named isoforms=3;
 CC Comment-Experimental confirmation may be lacking for some
 CC isoforms;
 CC Name=1;

CC Isoid=Q9UIF8-1; Sequence=Displayed;
 CC Name=2;
 CC Isoid=Q9UIF8-2; Sequence=VSP_000553;
 CC Note=Inferred from Ref.4;
 CC Name=3;
 CC Isoid=Q9UIF8-3; Sequence=VSP_000554;
 CC Note=Inferred from Ref.5;
 CC -1- TISSUE SPECIFICITY: Expressed at varying levels in several
 CC tissues, whereas a smaller transcript was expressed specifically
 CC in testis.
 CC -1- SIMILARITY: Belongs to the WAL family.
 CC -1- SIMILARITY: Contains 1 bromodomain.
 CC -1- SIMILARITY: Contains 1 DDT domain.
 CC -1- SIMILARITY: Contains 1 methyl-CpG-binding (MBD) domain.
 CC -1- SIMILARITY: Contains 1 PHD-type zinc finger.
 CC -1- CAUTION: Ref.5 sequence differs from that shown due to a
 CC frameshift in position 731.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AB032255; BAA89212.1; -;
 CC EMBL; AB040909; BAA96000.2; ALT_INIT.
 CC EMBL; AL834381; CAD39044.1; ALT_INIT.
 CC EMBL; BC012576; AAH12576.1; ALT_FRAME.
 CC EMBL; AK027612; BAB55231.1; ALT_INIT.
 CC EMBL; AL080173; CAB45759.1; -;
 CC PIR; T12495; T12495.
 CC HSP; Q9UIG0; I162.
 CC Genew; HGNC:963; BAZ2B.
 CC MIM; 605683; -;
 CC InterPro; IPR001487; Bromodomain.
 CC InterPro; IPR004022; DDT.
 CC InterPro; IPR001739; Methyl-CpG_bind.
 CC InterPro; IPR001965; Znf_PHD.
 CC Pfam; PF00439; Bromodomain; 1.
 CC Pfam; PF02791; DDT; 1.
 CC Pfam; PF01429; MBD; 1.
 CC Pfam; PF00628; PHD; 1.
 CC PRINTS; PR00503; BROMODOMAIN.
 CC PROSITE; PS00633; BROMODOMAIN_1; FALSE_NEG.
 CC PROSITE; PS50014; BROMODOMAIN_2; 1.
 CC PROSITE; PS50827; DDT; 1.
 CC PROSITE; PS50982; MBD; 1.
 CC PROSITE; PS01359; ZF_PHD_1; FALSE_NEG.
 CC PROSITE; PS50016; ZF_PHD_2; 1.
 CC Alternative splicing; Bromodomain; Coiled coil; DNA-binding;
 CC Nuclear protein; Transcription regulation; Zinc-finger.
 CC DOMAIN 543 614 MBD.
 CC DOMAIN 891 956 DDT.
 CC ZN_FING 1735 1785 PHD-type.
 CC DOMAIN 1881 1951 Bromodomain.
 CC DOMAIN 45 69 Ser-rich.
 CC DOMAIN 73 79 Poly-Glu.
 CC DOMAIN 399 470 Asp/Glu-rich (acidic).
 CC DOMAIN 641 672 Arg-rich.
 CC DOMAIN 687 865 Coiled coil (Potential).
 CC DOMAIN 706 875 Lys-rich.
 CC DOMAIN 1100 1143 Asp-rich.
 CC DOMAIN 1138 1179 Coiled coil (Potential).
 CC VARSPLIC 437 534 Missing (in isoform 2).
 CC VARSPLIC 593 626 Missing (in isoform 3).
 CC VARSPLIC 137 137 /FTid=VSP_000554.
 CC CONFLICT 226 226 S -> F (in Ref. 1).
 CC CONFLICT 627 627 S -> L (in Ref. 1).
 CC CONFLICT 722 722 G -> E (in Ref. 1).
 CC CONFLICT 722 722 E -> K (in Ref. 5).

FT	CONFLICT	790	799	RIKEMRRQ -> QKKKKKKKK (in Ref. 5).
FT	CONFLICT	1128	1128	K -> Q (in Ref. 1).
FT	CONFLICT	1183	1183	Q -> P (in Ref. 1).
FT	CONFLICT	1195	1195	Q -> R (in Ref. 1).
FT	CONFLICT	1453	1453	L -> S (in Ref. 3).
FT	CONFLICT	1838	1838	K -> Q (in Ref. 1).
SQ	SEQUENCE	1972	AA; 220709 MM; D4F105C927B91569 CRC64;	
Query Match				
Best Local Similarity 99.6%; Score 10097; DB 1; Length 1972;				
Matches 1965; Conservative 3; Mismatches 4; Indels 0; Gaps 0;				
Qy	1	MGQKSTSSGGNRKCNQEQSNQKQPLDARVVKIKDKPKPKKAMESSSNSDSSTGSSDTS	60	
Db	1	MGQKSTSSGGNRKCNQEQSNQKQPLDARVVKIKDKPKPKKAMESSSNSDSSTGSSDTS	60	
Qy	61	SEGLSSSDSDLEDEDEREEDSIESDDSDSESEAOHKNNOVLLHGI	120	
Db	61	SEGLSSSDSDLEDEDEREEDSIESDDSDSESEAOHKNNOVLLHGI	120	
Qy	121	TEKAQEKRIHOPLPLAFESQTHSFOSQKQVLSQQLPFIQSOAKKEESVKNKHTSVIQ	180	
Db	121	TEKAQEKRIHOPLPLAFESQTHSFOSQKQVLSQQLPFIQSOAKKEESVKNKHTSVIQ	180	
Qy	181	STGLSVNVKPLSLVNQAKKTYMKLIIVSPDVLKAGNKTSESSLLTSELRSKREQYKQ	240	
Db	181	STGLSVNVKPLSLVNQAKKTYMKLIIVSPDVLKAGNKTSESSLLTSELRSKREQYKQ	240	
Qy	241	APPSOLKQESSKSLKVIKIALSNPKATSSPAHPKOTLENNHNPFLTALLGNHQNG	300	
Db	241	APPSOLKQESSKSLKVIKIALSNPKATSSPAHPKOTLENNHNPFLTALLGNHQNG	300	
Qy	301	VIQSVIQEAPLALTTKTMQSKINENIAAASSTPFSSPVNLSTSGRRTPGNOTPMPASAS	360	
Db	301	VIQSVIQEAPLALTTKTMQSKINENIAAASSTPFSSPVNLSTSGRRTPGNOTPMPASAS	360	
Qy	361	PILHSQKKEKAVSNVNPVKTOHSHHPAKSLVEQPRGTDSDIPSKOSEDSENEDEEDE	420	
Db	361	PILHSQKKEKAVSNVNPVKTOHSHHPAKSLVEQPRGTDSDIPSKOSEDSENEDEEDE	420	
Qy	421	EEDEEDDEDDSDSQSDNSSDTSGSEEDDDKQDESDDTGEKTSMLNKTT	480	
Db	421	EEDEEDDEDDSDSQSDNSSDTSGSEEDDDKQDESDDTGEKTSMLNKTT	480	
Qy	481	SSVKSPSMSLTHSTPRNLHTAKPGSAPALCSQSPAFILGTSSTLTSSPHSGTSKR	540	
Db	481	SSVKSPSMSLTHSTPRNLHTAKPGSAPALCSQSPAFILGTSSTLTSSPHSGTSKR	540	
Qy	541	RRVTDRELRIPILEYGWQRETRIRNFGRLQGEVAYAPCGKKLRQYPEVIKILSRNGIM	600	
Db	541	RRVTDRELRIPILEYGWQRETRIRNFGRLQGEVAYAPCGKKLRQYPEVIKILSRNGIM	600	
Qy	601	DISRDNFSKIRVGDPIYARDGQEQWCLLKEEDVIPRIWEGRRGPPNPDRORA	660	
Db	601	DISRDNFSKIRVGDPIYARDGQEQWCLLKEEDVIPRIWEGRRGPPNPDRORA	660	
Qy	661	REESMRRRKGRPPNVGNAEFLDNADAKLLKLOAQETARQAQIKLLRKLQKQEOARVA	720	
Db	661	REESMRRRKGRPPNVGNAEFLDNADAKLLKLOAQETARQAQIKLLRKLQKQEOARVA	720	
Qy	721	KEAKKQQAIAAAEKKRQKEQIKIMKQOQEKIKRIQIIRMEKELRAAQILEAKKKKEEAA	780	
Db	721	KEAKKQQAIAAAEKKRQKEQIKIMKQOQEKIKRIQIIRMEKELRAAQILEAKKKKEEAA	780	
Qy	781	NAKLLAEKRIKEEMRRQQAALLKHQBRERRRRQHMLMKAMEARKKAAEERLKKQEKRD	840	
Db	781	NAKLLAEKRIKEEMRRQQAALLKHQBRERRRRQHMLMKAMEARKKAAEERLKKQEKRD	840	
Qy	841	EKRLNKERKLQRRLLEMAKELKPNEDMCLADQKPLPELPRIPGLVLSGSTSDCLMW	900	
Db	841	EKRLNKERKLQRRLLEMAKELKPNEDMCLADQKPLPELPRIPGLVLSGSTSDCLMW	900	
Qy	901	VQFLRNFGLVGFVDNIDVPNLSVLQEGLLNIGSMGEVQDILLVRLLSAAVCDPLITGY	960	

Db	901	VQFLRNFGLVGFVDNIDVPNLSVLQEGLLNIGSMGEVQDILLVRLLSAAVCDPLITGY	960	
Qy	961	KAKTALGEBHLNVGNVRDNVSEILQIFMEAHCGOTELTESLTKTAFQAHTPAQKASVLAF	1020	
Db	961	KAKTALGEBHLNVGNVRDNVSEILQIFMEAHCGOTELTESLTKTAFQAHTPAQKASVLAF	1020	
Qy	1021	LINELACSKSVSEIDKNIDYMSNLRDKWVVEGKLRLRIHAKTKGRDTSGGIDIGE	1080	
Db	1021	LINELACSKSVSEIDKNIDYMSNLRDKWVVEGKLRLRIHAKTKGRDTSGGIDIGE	1080	
Qy	1081	BOHPLGTPTPKERRRKGDSYDDDDDDDDDDDDDEDEDEDEDEDEDEDEDEDEDEDEDE	1140	
Db	1081	BOHPLGTPTPKERRRKGDSYDDDDDDDDDDDDDEDEDEDEDEDEDEDEDEDEDEDEDE	1140	
Qy	1141	EGDOAASVEELEKQIEKLSKQOSQYRRKLFDAHSLSRSMFGPDYRRYRYWTLPRCGGIF	1200	
Db	1141	EGDOAASVEELEKQIEKLSKQOSQYRRKLFDAHSLSRSMFGPDYRRYRYWTLPRCGGIF	1200	
Qy	1201	VEGMESGEGLEBEIAKEREKLAESVQIKEEMFETSGDSLNCSTNDHCEQEDLKEKONT	1260	
Db	1201	VEGMESGEGLEBEIAKEREKLAESVQIKEEMFETSGDSLNCSTNDHCEQEDLKEKONT	1260	
Qy	1261	NLFLOKPGSFSLKLEVAOKPPESEVMTPKPNAGANGCTLSYQNSGKHSLSGVQSTAT	1320	
Db	1261	NLFLOKPGSFSLKLEVAOKPPESEVMTPKPNAGANGCTLSYQNSGKHSLSGVQSTAT	1320	
Qy	1321	QSNVEKADSNLNTGSGPGKFSYPLPNDQLLTKLTETKNQWFSLLPRTPCDDTSLTHA	1380	
Db	1321	QSNVEKADSNLNTGSGPGKFSYPLPNDQLLTKLTETKNQWFSLLPRTPCDDTSLTHA	1380	
Qy	1381	DMSTASLVTPOSPSPSPSPPTPAPLGSSAQNPVGLNPFALSPLOVKGVSMMGLQFCGW	1440	
Db	1381	DMSTASLVTPOSPSPSPSPPTPAPLGSSAQNPVGLNPFALSPLOVKGVSMMGLQFCGW	1440	
Qy	1441	PTGVVTSNIPPTLSVPSLGSGLGSEGNGSFLTSNVASSKSESVPQNEKATSQAQAAV	1500	
Db	1441	PTGVVTSNIPPTLSVPSLGSGLGSEGNGSFLTSNVASSKSESVPQNEKATSQAQAAV	1500	
Qy	1501	EVAKVPDPPSPKPIPEEMQFCGWRIIDPEDLAKLLKVLHLRGIRKALQKQIKQLHDYIT	1560	
Db	1501	EVAKVPDPPSPKPIPEEMQFCGWRIIDPEDLAKLLKVLHLRGIRKALQKQIKQLHDYIT	1560	
Qy	1561	QACLKNDKVAIIELNENEENQVTRDIVENWVSEEQAMEMDLSVLQOVEDLERRVASASIQ	1620	
Db	1561	QACLKNDKVAIIELNENEENQVTRDIVENWVSEEQAMEMDLSVLQOVEDLERRVASASIQ	1620	
Qy	1621	VKGWMCPEPASERDLVYFEHKSFTKLCHEHDEFTGEDESSAHALERKSNPLDIAVTR	1680	
Db	1621	VKGWMCPEPASERDLVYFEHKSFTKLCHEHDEFTGEDESSAHALERKSNPLDIAVTR	1680	
Qy	1681	LADLERNTERIERIEDIAPGLVWRRALSEARSAQAVALCICLOLOKSIKAKSOTLKIKLHVKG	1740	
Db	1681	LADLERNTERIERIEDIAPGLVWRRALSEARSAQAVALCICLOLOKSIKAKSOTLKIKLHVKG	1740	
Qy	1741	CRKGDNEELLCDGCKCHTYCHRPKITIIPDGDFPCACIACASQTLKIKLHVKG	1800	
Db	1741	CRKGDNEELLCDGCKCHTYCHRPKITIIPDGDFPCACIACASQTLKIKLHVKG	1800	
Qy	1801	KKTNESKKGKVTULTGTDEDEDSASTSSSLKRGKNDLQKRWMEENTSINLSKQESFTSVK	1860	
Db	1801	KKTNESKKGKVTULTGTDEDEDSASTSSSLKRGKNDLQKRWMEENTSINLSKQESFTSVK	1860	
Qy	1861	KPKEDDSKDLALCSMLTETHEDEAMPPLPVNLKLVPGYKVKIKKPMDESTIREKLSS	1920	
Db	1861	KPKEDDSKDLALCSMLTETHEDEAMPPLPVNLKLVPGYKVKIKKPMDESTIREKLSS	1920	
Qy	1921	GOYFNLETFAIDVRLVFNDCETFNEDDDSDIGRAGHNMRKYFEKKWTDTFKVS	1972	
Db	1921	GOYFNLETFAIDVRLVFNDCETFNEDDDSDIGRAGHNMRKYFEKKWTDTFKVS	1972	


```
QY 1066 KTGKRDTSGGIDLGEEHPLGTPTGCRKRRKGGSDYDDDDDDSDSDQDDEDEBDK 1125
Db 1185 KTGKRDATGGGVEPEHSLTPTGCRKRRKGGSDYDDDDSDSDQDDEDEBDK 1244
QY 1126 EDQKGTDTICEDDEGQAAVELEKQIEKLSKQSQYRRKLFDAHSLSRSLVWFGPDR 1185
Db 1245 EDKKGKAEVEDEDDGQDTVSVELEKQIEKLTQSQYRRKLFDAHSLSRSLVWFGQDR 1304
QY 1186 YRRYRWWLPCCGIFVEGMEGEGLEETAKEREKLKKAESVQIKKEEMFETSGDSLNCNT 1245
Db 1305 YRRYRWWLPCCGIFVEGMEGEGLEETAKEREKLKKAESVQIKKEEMFETSGDSLNCNT 1364
QY 1246 DHCEQKDLKEDNTNLFQPGSFSLKSLLEVAKMPPESEVWTPKKNAG-ANGCTLSY 1304
Db 1365 TRCEQKDLKEDNTNLFQPGSFSLKSLLEVAKMPPESEVWTPKKNAG-ANGCTPSY 1424
QY 1305 QNSGHSLSGVSQSTATQSNVKAENNLFTNGSSGPKGYSP--LPNDOLLKTLTEKQRQ 1362
Db 1425 QNTSONSLCSLQPSVQSSEKSDSNLFSPTASGTGKFPYSPPLIPSDLLKTLTEKQRQ 1484
QY 1363 WFSLLPRTPCDDTSLTHADMTASLVTPSQSPSPSPPTAPL--GSSAQNPVGLNPPA 1420
Db 1485 WFSLLPRTPCDDTSLTHADMTASLVTPSQSPSPSPPTAPL--GSSAQNPVGLNPPA 1544
QY 1421 LSPL-QVKGVSMMGLQFCGHPGTGVTGNIPTLSVPSGLSLGSEGNFSFLTNSVAS 1479
Db 1545 MPPLQOMKPLPVNMGLOQFCGHPGTGVTGNIPTLSVPSGLSLGSEGNFSFLTNSVAS 1604
QY 1480 SKSPSPVQNEKATSAQPAAEVAKPVPDPSPKPIPEEMQGWRIIDPDLKALLKVLH 1539
Db 1605 SKSPSPVQNEKATSAQPAAEVAKPVPDPSPKPIPEEMQGWRIIDPDLKALLKVLH 1664
QY 1540 LRGIREKALQIQKHLDTIYQACLKNDKVAIIELENEENQVTRDVENNVSVEEQAMEV 1599
Db 1665 LRGIREKALQIQKHLDTIYQACLKNDKVAIIELENEENQVTRDVENNVSVEEQAMEV 1724
QY 1600 DLSVLQVEDLERRVASASLQVGMCPPEASEREDLVYFHKFTLCKEHDGEFTGED 1659
Db 1725 DLSVLQVEDLERRVASASLQVGMCPPEASEREDLVYFHKFTLCKEHDGEFTGED 1784
QY 1660 ESSAHALERKSDNPLDIATVRLADLERNIER----- 1691
Db 1785 EGSTSSLERKNDNPLDIATVRLADLERNIERYLKSPSTTIQIKLDNVGTVTVAPAPPS 1844
QY 1692 -----IEEDIAPGLRVRRALSEARSAAQVALCICQLOKSIWEKSIKMYVQICRKG 1744
Db 1845 ISGDGDTGTEEDIAFGLRVRRALSEARSAAQVALCICQLOKSIWEKSIKMYVQICRKG 1904
QY 1745 DNEELLCLDCGCDKGCHTYCHRPKITTIPTDGDWFCPACIAKASGQTLKIKLHVKGKTN 1804
Db 1905 DNEELLCLDCGCDKGCHTYCHRPKITTIPTDGDWFCPACIAKASGQTLKIKLHVKGKTN 1964
QY 1805 ESKGKKVTLTCDTEDEDSATSSSLKGNKDLQKRNKMEENTSINLSQESFTSVKPKR 1864
Db 1965 ESKGKKVTLTCDTEDEDSATSSSLKGNKDLQKRNKMEENTSINLSQESFTSVKPKR 2022
QY 1865 DSKDLALCSMLTMEHEDAWPFLPVNKLVPYKVKIKKPMDFSTIREKLSGGOYP 1924
Db 2023 DSKDLALCSMLTMEHEDAWPFLPVNKLVPYKVKIKKPMDFSTIREKLSGGOYP 2082
QY 1925 NLETALDVLRFVDFNCETFNEDSDIGRAGHNMRYPEKKWTDTPKVS 1972
Db 2083 NLETALDVLRFVDFNCETFNEDSDIGRAGHNMRYPEKKWTDTPKVS 2130

RESULT 3
Q69218 ID Q69218 PRELIMINARY; PRT; 1369 AA.
AC Q69218:
DT 25-OCT-2004 (TrenBRel. 28, Created)
DT 25-OCT-2004 (TrenBRel. 28, Last sequence update)
DT 25-OCT-2004 (TrenBRel. 28, Last annotation update)
DE MKIAA1476 protein (Fragment).
```

```
GN Name=MKIAA1476;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Okazaki N., Kikuno R.F., Ohara R., Inamoto S., Koseki H., Hiraoka S.,
RA Soga Y., Seino S., Nishimura M., Kaisho T., Hoshino K., Kitamura H.,
RA Nagase T., Ohara O., Koga H.;
RT "Prediction of the Coding Sequences of Mouse Homologues of KIAA Gene:
RT IV. The Complete Nucleotide Sequences of 500 Mouse KIAA-Homologous
RT cDNAs Identified by Screening of Terminal Sequences of cDNA Clones
RT Randomly Sampled from Size-Fractionated Libraries.";
RL DNA Res. 11:205-218(2004).
DR EMBL; AK173180; "BAD32458-1";
DR GO; GO:0003677; F:DNA binding; IEA.
DR InterPro; IPR004022; DDT.
DR InterPro; IPR001739; Methyl-CpG_bind.
DR Pfam; PF02791; DDT; 1.
DR Pfam; PF01429; MBD; 1.
DR SMART; SM00571; DDT; 1.
DR SMART; SM00391; MBD; 1.
DR PROSITE; PSS0827; DDT; 1.
FT NON TER 1
FT NON TER 1369
SQ SEQUENCE 1369 AA; 150394 MW; 9ACFC55329361A33 CRC64;

Query Match 52.0%; Score 5274; DB 2; Length 1369;
Best Local Similarity 85.6%; Pred. No. 1.8e-191;
Matches 1042; Conservative 55; Mismatches 54; Indels 66; Gaps 8;

QY 1 MGQKTSSTSGGKRNKQEQSKNQPLDARVDKIKDKPKKAMSSSSSDSDSGTSSDTS 60
Db 215 MGQKTSSTSGGKRNKQEQSKNQPLDARVDKIKDKPKKAMSSSSSDSDSGTSSDTS 274
QY 61 SEG1SSSDSDLEDEDEDEDSIESEDDSDSDSESAQHKNQVLLHGISDPKADGKA 120
Db 275 SEG1SSSDSDLEDEDEDEDSIESEDDSDSDSESAQHKNQVLLHGISDPKADGKA 333
QY 121 TEKAQEKRIHQPLAPAFESOTH-SFOOQKOPQVLSOOLPEIFOSSQAKESVNHKTSVI 179
Db 334 TEKAQEKRIHQPLAPAFESOTH-SFOOQKOPQVLSOOLPEIFOSSQAKESVNHKTSVI 393
QY 180 QSTGLSVNVPKPLSLVNOAKKTYMKLVPSPDVLKAGNKNTSESSLLTSELRSKREYK 239
Db 394 QSTGLSVNVPKPLSLVNOAKKTYMKLVPSPDVLKAGNKNTSESSLLTSELRSKREYK 453
QY 240 QAFPSQLKQSESSKSLKVAALSNPKATSSSPAPHPKQTLNHNHPNPLTNALLGNHPN 299
Db 454 QAFPSQLKQSESSKSLKVAALSNPKATSSSPAPHPKQTLNHNHPNPLTNALLGNHPN 513
QY 300 GVIOSTVQEAFLATTTKQSKINENIAAASSTPSSPVNLSTSGRTPGNQTPVMPBSA 359
Db 514 GVIOSTVQEAFLATTTKQSKINENIAAASSTPSSPVNLSTSGRTPGNQTPVMPBSA 571
QY 360 SPILHSSGKEKAVSNVNPVKTQHHSHP--AKSLVEQFRGTGSDIPSSKDESDNEDEE 416
Db 572 SPILHSSGKEKAVSNVNPVKTQHHSHP--AKSLVEQFRGTGSDIPSSKDESDNEDEE 631
QY 417 EDDDEEEDDEDDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSD 476
Db 632 EDDDEEEDDEDDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSD 689
QY 477 NKTTSVKSPSMSLTGHTSTPRNLHIAKAPGAPALCSESQSPALFCTSSSTLTSSPHSG 536
Db 690 NKTTSVKSPSMSLTGHTSTPRNLHIAKAPGAPALCSESQSPALFCTSSSTLTSSPHSG 748
QY 537 TSKRRVTDRELRIPLEYGMQRETRIRNFGRIGQEVAYVAYPCGKKLRQVPEVIKLSR 596
Db 749 TSKRRVTDRELRIPLEYGMQRETRIRNFGRIGQEVAYVAYPCGKKLRQVPEVIKLSR 804
QY 597 NGIMDISRDNFSFAKIRVGDYFYEARDGPQBMQWCLLKEEDVIPRIRAMEGRGRPPND 656
```



```
Db 805 -----GQWQCKLKEEDVIRIRAMDGRGRPNPD 834
Qy 657 RQAREBSMRKRGRPPNVGNABFLNADAKLRKLAQAEIARQAAQIKLRLKLOQEQ 716
Db 835 RPRAREBSMRKRGRPPNVGSABFLDNTDAKLRKLAQAEIARQAAQIKLRLKLOQEQ 894
Qy 717 ARVAKEAKKQAAIWAEEKRRKQEIIMKQOEKI KRIQIIRMEKELRAQOILBAKXKK 776
Db 895 ARVAKEAKKQAAIWAEEKRRKQEKMIKQOEKI KRIQIIRMEKELRAQOILB 948
Qy 777 EEAANAKLLEAKRIKEKEMRRQAAVLKHQERERRRRCHMLMKAMEARKKAEERLKQ 836
Db 949 -----EKELRRQAVLLKHQERERRRRQHVMKAMEARKKAEERLKQ 992
Qy 837 EKREKELNKERKLEQRRELEMAKELKKNEDNCLADQKPLPELPRI PGLVLVSGTFS 896
Db 993 EKREKELNKERKLEQRRELEMAKELKPKEDNCLADQKPLPEWPRI PGLVLVSGTFS 1052
Qy 897 CLMVVQFLRNFGKVLGFDVNI DVNLSVLQGLLNI GDSMGEVDLLVRLLSAAVCDPGL 956
Db 1053 CLMVVQFLRNFGKVLGFDVNI DVNLSVLQGLLNI GDSMGEVDLLVRLLSAAVCDPGL 1112
Qy 957 ITGYKAKTALGEHLNAGVNRDNVSEILQIFMEAHCGQTELTESLTKTKAFOAHTPAQKAS 1016
Db 1113 ITGYKAKTALGEHLNAGVNRDNVSEILQIFMEAHCGQTELTESLTKTKAFOAHTPAQKAS 1172
Qy 1017 VLAFLINELACSKSVSEIDKNIIDYMSNLRDRKQWVVGKLRKLRIRIIHAKTKGRDRTSGGI 1076
Db 1173 ILAFLVNLACSKSVSEIDKNIIDYMSNLRDRKQWVVGKLRKLRIRIIHAKTKGRDRTSGGI 1232
Qy 1077 DLGEEQHPGTPTGKRKRKRGSDYDDDDDDDDQDQDDEDEDKEDKGGKTTDIC 1136
Db 1233 DLGEEQHPGTPTGKRKRKRGSDYDDDDDDDDQDQDDEDEDKEDKGGKTTDIC 1292
Qy 1137 EDEDEGQAAVELEKQIEKLSKQSQYRKRKLPDASHSLRSVMFGPDYRRRRYWIILPRC 1196
Db 1293 EDEDEGQTAASVELEKQIEKLSKQSQYRKRKLPDASHSLRSVMFGPDYRRRRYWIILPQC 1352
Qy 1197 GGIFVEGMESGEGLEEI 1213
Db 1353 GGIFVEGMESGEGLEEI 1369

RESULT 4
Q8NC87 ID Q8NC87 PRELIMINARY; PRT; 796 AA.
AC Q8NC87;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein FLJ90414.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCHI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Isegai T., Ota T., Nishikawa T., Hayashi K., Otsuki T., Sugiyama T.,
RA Suzuki Y., Nagai K., Sugano S., Ishii S., Kawai-Hio Y., Saito K.,
RA Yamamoto J., Wakamatsu A., Nakamura Y., Kojima S., Nagahari K.,
RA Masuho Y., Ono T., Okano K., Yoshikawa Y., Aotsuka S., Sasaki N.,
RA Hattori A., Okumura K., Iwayanagi T., Ninomiya K.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK074895; BAC11274.1; -.
DR HSP; Q9UIG0; 1F62.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR011011; FIVE_PHD_Znf.
DR InterPro; IPR001965; Znf_PHD.
DR Pfam; PF00628; PHD; 1.
DR SMART; SM00249; PHD; 1.
DR PROSITE; PS50016; 2F_PHD_2; 1.
```

```
FT NON TER 796 796
SQ SEQUENCE 796 AA; 88462 MW; 2EDB8BCAF9720C1E CRC64;

Query Match 41.0%; Score 4153; DB 2; Length 796;
Best local Similarity 99.5%; Pred. No. 2.4e-149;
Matches 792; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1042 MSNLRDRKQWVEGKLRKLRIRIIHAKTKGRDTSGGIDLGEBOHPLGTPTGKRKRKRGKGD 1101
Db 1 MSNLRDRKQWVEGKLRKLRIRIIHAKTKGRDTSGGIDLGEBOHPLGTPTGKRKRKRGKGD 60
Qy 1102 DYDDDDDDDDQDQDDEDEDKEDKGGKTTDICEDEDEGQAAASVEELEKQIEKLSKQ 1161
Db 61 DYDDDDDDDDQDQDDEDEDKEDKGGKTTDICEDEDEGQAAASVEELEKQIEKLSKQ 120
Qy 1162 OSQYRKRKLPDASHSLRSVMFGPDYRRRRYWIILPRCGGIFVEGMESGEGLEIAKERK 1221
Db 121 OSQYRKRKLPDASHSLRSVMFGPDYRRRRYWIILPRCGGIFVEGMESGEGLEIAKERK 180
Qy 1222 KAESVQIKKEEMFETSGDSLNCSTNDHCEQEDLKEKDNNTLFLQKPGSFSLKLELVAK 1281
Db 181 KAESVQIKKEEMFETSGDSLNCSTNDHCEQEDLKEKDNNTLFLQKPGSFSLKLELVAK 240
Qy 1282 MPPESEVMTPKPNAGANGCTLSYQNSGKHSLSGVSQSTATQSNVEKADSNNTLNTGSSGPG 1341
Db 241 MPPESEVMTPKPNAGANGCTLSYQNSGKHSLSGVSQSTATQSNVEKADSNNTLNTGSSGPG 300
Qy 1342 KFYSPLPNDQLLTKLTEKNRQWFSLLPRTPCDDTSLTHADMSTASLTVPQSPPSKSP 1401
Db 301 KFYSPLPNDQLLTKLTEKNRQWFSLLPRTPCDDTSLTHADMSTASLTVPQSPPSKSP 360
Qy 1402 TPAPLASSAONPVGLNPFALSPLOVKGVSMMGLQFCGWPTGVVTSNIPTTSSVPSLSG 1461
Db 361 TPAPLASSAONPVGLNPFALSPLOVKGVSMMGLQFCGWPTGVVTSNIPTTSSVPSLSG 420
Qy 1462 LGLSEGNGSFLTSNVASSKSESPVPQNEKATSAQPAAVEAVKPVDFPSPKPIPEEMQFG 1521
Db 421 LGLSEGNGSFLTSNVASSKSESPVPQNEKATSAQPAAVEAVKPVDFPSPKPIPEEMQFG 480
Qy 1522 WRIIDPEDLKALLKVLHLRGIREKALQKQIQHLDYITQACLKNKDVAIIELNENEENQ 1581
Db 481 WRIIDPEDLKALLKVLHLRGIREKALQKQIQHLDYITQACLKNKDVAIIELNENEENQ 540
Qy 1582 VTRDIVENWSVEQAMEMDLSVLQOVEDLERRVASASLOVKGWMCPPASEREDLVTFEH 1641
Db 541 VTRDIVENWSVEQAMEMDLSVLQOVEDLERRVASASLOVKGWMCPPASEREDLVTFEH 600
Qy 1642 KSFTKLCKEHGDGFTGEDESSAHLERKSDNPLDIAVTRLADLERNIERRIEEDIAFGLR 1701
Db 601 KSFTKLCKEHGDGFTGEDESSAHLERKSDNPLDIAVTRLADLERNIERRIEEDIAFGLR 660
Qy 1702 VRRALSEARSAQAVALCIIQOLQKSIWAKSINWKYCOICRKGDNBELLLLCDGCDKGCH 1761
Db 661 VRRALSEARSAQAVALCIIQOLQKSIWAKSINWKYCOICRKGDNBELLLLCDGCDKGCH 720
Qy 1762 TYCHRPKITTPDGMWFCPACIAKAGQTLKIKLHVKGKKTNESKKGKVTLTGDTDE 1821
Db 721 TYCHRPKITTPDGMWFCPACIAKAGQTLKIKLHVKGKKTNESKKGKVTLTGDTDE 780
Qy 1822 DSASTSSSLKRGKNDL 1837
Db 781 DSASTSSSLKRGKNDL 796

RESULT 5
Q8COK4 ID Q8COK4 PRELIMINARY; PRT; 861 AA.
AC Q8COK4;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Mus musculus adult male thymus cDNA, RIKEN full-length enriched
DE library, clone:5830435C13 product:BROMODOMAIN ADJACENT TO ZINC FINGER
```



```

Db 1492 ROLPAFO---EGIMWSWSPKRYETDGLAVLQWVELEQRMVMSDLQIRGWTCPSPDSTRE 1548
Qy 1635 DLVTFEHSFKYCKLCKEHGDEFTGDESSAHALERKSDNPLDIAVTRLDLERNTER--- 1691
Db 1549 DLAYCEHLSDS---QEDITWRGRGREGI-APQKTTNPLDLAVWRLAALQONVRRYL 1603
Qy 1692 -----IIEDIAPGLRVRRRLSEARSAAQVALCIQ 1722
Db 1604 EPLWPTHEVLEKALLSTPNGAPETTTIEISYEITPRIVVRQTLERCSRSAQVCLCLGQ 1663
Qy 1723 LOKSIAMEKSTMKVYCOICRGDNEELLLCGDKGCHTYCHRPKITTIPDGDMFCPCAC 1782
Db 1664 LERSIAEKSVNRYTCLVCRGDNDFLLCGDGRGCHYCHRPKMEAVPEGDMFCVC 1723
Qy 1783 IAKASGQTLKIKLVHVGKKTNE---SKKGKKVLTGTDEDSASTSSSLKRGKNDKOL 1839
Db 1724 LAQ-----QVEGEFTQKGFPRGQK-----RKSGYSLNFSGEGDGR 1760
Qy 1840 RKM-----EENTSNILSKQESTSVKPK-----RDSKDLALCSMLTETMETHEDAWPFL 1891
Db 1761 RVLRLRGESPAAGPRYSSEGLSPSKRRRLSMRHHSDLTFCETILMEMESHDAWPFLE 1820
Qy 1892 PVNLKLVPGYKVKIKWDFSTIBELSSGOPYNLETFALDVLRLVFNFCENTFEDDDIG 1951
Db 1821 PVNRLVSGVYRIIKNPMDFMTMRLLRGYTSSEFEPAADALLVFNFCNTFEDDSEVG 1880
Qy 1952 KAGHMRKRYFEKKTWDTFK 1970
Db 1881 KAGHMRFFESRWEFQ 1899

RESULT 7
BA2A_HUMAN
ID_BA2A_HUMAN STANDARD; PRT; 1878 AA.
AC Q9UIF9; O00536; O15030; Q96H26;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Bromodomain adjacent to zinc finger domain 2A (Transcription
DE termination factor-1 interacting protein 5) (TTF-I interacting protein
DE 5) (TTF5) (hWALp3).
GN Name=BAZ2A; Synonyms=K1AA0314, TIP5;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RX MEDLINE=20130112; PubMed=10662543; DOI=10.1006/geno.1999.6071;
RA Jones M.H., Hamana N., Nezu J.-I., Shimane M.;
RT "A novel family of bromodomain genes.";
RL Genomics 63:40-45(2000).
RN [2]
RP SEQUENCE OF 332-738 FROM N.A.
RC TISSUE=Lung;
RA Jansa P., Grummt I.;
RT Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 639-1878 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=97349984; PubMed=9205841;
RA Nagase T., Ishikawa K.-I., Nakajima D., Ohira M., Seki N.,
RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. VII.
RT The complete sequences of 100 new cDNA clones from brain which can
RL code for large proteins in vitro.";
RL DNA Res. 4:141-150(1997).
RN [4]
RP SEQUENCE OF 1038-1878 FROM N.A.
RC TISSUE=Lymph;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;

```

```

RA Strauberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh P.,
RA Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Schetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahay J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butcherfield Y.S.N., Kryzyski M.I., Skalska U., Smalls D.E.,
RA Sutterch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -!- FUNCTION: May play a role in transcriptional regulation
CC interacting with ISWI. May serve a specific role in maintaining or
CC altering the chromatin structure of the rDNA locus (By
CC similarity).
CC -!- SUBUNIT: Together with ISWI/SNF2h, it forms a complex termed NORC
CC (nucleolar remodeling complex).
CC -!- SUBCELLULAR LOCATION: Nuclear. Colocalizes with the basal RNA
CC polymerase I transcription factor UBF in the nucleolus.
CC -!- TISSUE SPECIFICITY: Expressed at moderate levels in most tissues
CC analyzed, including heart, brain, placenta, lung, skeletal muscle,
CC kidney and pancreas.
CC -!- SIMILARITY: Belongs to the WAL family.
CC -!- SIMILARITY: Contains 4 A.T hook DNA-binding repeats.
CC -!- SIMILARITY: Contains 1 bromodomain.
CC -!- SIMILARITY: Contains 1 DDT domain.
CC -!- SIMILARITY: Contains 1 methyl-CpG-binding (MBD) domain.
CC -!- SIMILARITY: Contains 1 PHD-type zinc finger.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB032254; BAA89211.1; -
DR EMBL; AF000422; AAB60864.1; -
DR EMBL; AB002312; BAA20773.1; -
DR EMBL; BC008965; AAB08965.2; -
DR HGSP; Q9UIG0; 1F62.
DR Genew; HGNC:962; BAZ2A.
DR MIM; 605682; -
DR GO; GO:0005731; C:nucleolus organizer complex; NAS.
DR GO; GO:0003677; F:DNA binding; NAS.
DR GO; GO:0030528; F:transcription regulator activity; NAS.
DR GO; GO:0006338; P:chromatin remodeling; NAS.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; NAS.
DR InterPro; IPR000637; A-T hook.
DR InterPro; IPR001487; Bromodomain.
DR InterPro; IPR004022; DDT.
DR InterPro; IPR001739; Methyl-CpG_bind.
DR Pfam; PF02178; AT_hook; 4.
DR Pfam; PF00439; Bromodomain; 1.
DR Pfam; PF02791; DDT; 1.
DR Pfam; PF01429; MBD; 1.
DR Pfam; PF00628; PHD; 1.
DR PRINTS; PR00503; BROMODOMAIN.
DR SMART; SM00384; AT_hook; 4.
DR SMART; SM00297; BROMO; 1.
DR SMART; SM00571; DDT; 1.
DR SMART; SM00391; MBD; 1.

```


Db 1219 GFLEQEGSPLSGQS-QHDLs--QSAFLSWLSQTSQSHSSLSVSVL--TPDSSQKGL-DP 1272
 Qy 1347 LPND-----OLKLTLEKRNQFSLRLPTPCDDTSLTHADMSTASIVTTPSQPPSKS- 1398
 Db 1273 APSQPPPEPEDEAESDPQALWFENISAQPCN-----AAATP---PPAVSE 1317
 Qy 1399 --PPTPAPLGSS--AQNVGNGNFPALSPLOQGVSNMGL---QFCWPTGVVTSNIPF 1451
 Db 1318 DOPTSPQOLASSKPMNRPSAANP--CSPQVF--SSTPLAGLAPKRRAGDPG----- 1365
 Qy 1452 TLSVPSLGSLGGLSEGNSELSNVASSSKSPVPQNEKATSAQPAAVEVAKPVDPPSP 1511
 Db 1366 --EMPQSTGLGQPKRCR-----PPSKP---FKQNEQRYLTQLTA----- 1401
 Qy 1512 KPIPEWQFGWRIIDPEDLKALIKVLHLRGIREKALQKQIQKHLDYITQACLKKNQVAI 1571
 Db 1402 QVPPEMCGWGWIRDPEDMLKALHPRGIREKALHKLHNRDFLEQVCLRPSADPI 1461
 Qy 1572 IELNENENQTRDIVENWSVEBQAMENDLSVLQVEDLRRVASASLQVKGWMCPEPAS 1631
 Db 1462 FEPRLQPAFQ---EGIMSWSPKPYETDYLAVLQWVELEQRVIMSDLIQIRGWTCPSPDS 1518
 Qy 1632 EREDLVFPHKSFYKLCCKEHDGEFTGEDESSAHALERKSDNPLDIATVRLADLERNIERR 1691
 Db 1519 TRELAYCEHLSDS---QEDITWRGRGREGI-APQKRTTNPLDLAVNRLAALBQNVERR 1573
 Qy 1692 -----IREDIAPGLRVWRRLSEARSAAQVALC 1719
 Db 1574 YLREPLMPTHEVLEKALLSTPNAGPEGTTVEISYEITPRVRVQTLERCESAAQVCLC 1633
 Qy 1720 IQQLQKSIAMKSKYKVKYQCRKGDNEILLCCDGGCHYCHRPKITIPDGMWFC 1779
 Db 1634 LGQLERSIAWKSVMKVTCLVCRKGDNDFFLLCCDGGCHYCHRPKMEAVEPGEWFC 1693
 Qy 1780 PACTAKASGQTLKIKLHVKKKTNE---SKGKKVTLTGTEDESDASTSSSLKRGKWD 1836
 Db 1694 TVCUAQ-----QVEGEFTQKPGFPRGQK-----RKSGYSLNFSEGD 1730
 Qy 1837 LQKRM-----BENTSINLSKQESFTSVKPKP-----RDSKDLALCSMLITMETHEDAWP 1888
 Db 1731 GRRRVLRLGREGSPAAGRYSEGLSPKRRSLNRHNSDLTCEIILMESHDAWP 1790
 Qy 1889 PLLPNVNLVPGYKVKIKKPMDFSTIREKLSGGQYVNLLETALDVLVDFDNCFTFNEDDS 1948
 Db 1791 FLEVPNPLVSGYRIIKNPMDFSTIREKLRGGYTSSEEPFAADALLVDFDNCFTFNEDDS 1850
 Qy 1949 DIGRAGHNRKYFEKKWTDTPK 1970
 Db 1851 EVGKAGHMRFFESRMEEFYQ 1872

RESULT 8

BA2A_MOUSE STANDARD; PRT; 1850 AA.
 ID_BA2A_MOUSE
 AC Q91YES;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 05-JUL-2004 (Rel. 44, Last annotation update)
 DE Bromodomain adjacent to zinc finger domain 2A (Transcription
 DE termination factor-1 interacting protein 5) (TF1F-1 interacting protein
 DE 5) (Tip5).
 GN Name=Baz2a; Synonyms=Tip5;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RZ SEQUENCE FROM N.A.
 RP MEDLINE=21423468; PubMed=11532953; DOI=10.1093/emboj/20.17.4892;
 RA Stroher R., Nemeth A., Jansa P., Hoffmann-Rohrer U., Santoro R.,
 RA Laengst G., Grummt I.;
 RT "NORC-a novel member of mammalian ISWI-containing chromatin remodeling

machines.";
 EMBL J. 20:4992-4900(2001).
 CC -!- FUNCTION: May play a role in transcriptional regulation
 CC interacting with ISWI. May serve a specific role in maintaining or
 CC altering the chromatin structure of the rDNA locus.
 CC -!- SUBUNIT: Together with ISWI/SNF2h, it forms a complex termed NORC
 CC (nucleolar remodeling complex).
 CC -!- SUBCELLULAR LOCATION: Nuclear. Colocalizes with the basal RNA
 CC polymerase I transcription factor UBF in the nucleolus.
 CC -!- SIMILARITY: Belongs to the WAL family.
 CC -!- SIMILARITY: Contains 4 A-T hook DNA-binding repeats.
 CC -!- SIMILARITY: Contains 1 bromodomain.
 CC -!- SIMILARITY: Contains 1 DDT domain.
 CC -!- SIMILARITY: Contains 1 methyl-CpG-binding (MBD) domain.
 CC -!- SIMILARITY: Contains 1 PHD-type zinc finger.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AJ309544; CAC69992.1; -.
 CC HSSP; Q9UG10; 1F62.
 DR MGD; MGI:2151152; Baz2a.
 DR GO; GO:0005731; C:nucleolus organizer complex; ISS.
 DR GO; GO:0003677; F:DNA binding; ISS.
 DR GO; GO:0030528; F:transcription regulator activity; ISS.
 DR GO; GO:0006338; P:chromatin remodeling; ISS.
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; ISS.
 DR InterPro; IPR000637; A-T hook.
 DR InterPro; IPR001487; Bromodomain.
 DR InterPro; IPR004022; DDT.
 DR InterPro; IPR001739; Methyl-CpG_bind.
 DR InterPro; IPR001965; Znfn_PHD.
 DR Pfam; PF02178; AT_hook; 4.
 DR Pfam; PF02791; DDT; 1.
 DR Pfam; PF01429; MBD; 1.
 DR Pfam; PF00628; PHD; 1.
 DR PRINTS; PR00503; BROMODOMAIN.
 DR SMART; SM00384; AT_hook; 4.
 DR SMART; SM00297; BROMO; 1.
 DR SMART; SM00571; DDT; 1.
 DR SMART; SM00391; MED; 1.
 DR SMART; SM00249; PHD; 1.
 DR PROSITE; PS00633; BROMODOMAIN_1; FALSE_NEG.
 DR PROSITE; PS00114; BROMODOMAIN_2; 1.
 DR PROSITE; PS50827; DDT; 1.
 DR PROSITE; PS50982; MBD; 1.
 DR PROSITE; PS01359; ZF_PHD_1; FALSE_NEG.
 DR PROSITE; PS50016; ZF_PHD_2; 1.
 KW Bromodomain; Coiled coil; DNA-binding; Nuclear protein; Repeat;
 KW Transcription regulation; Zinc-finger.
 FT DOMAIN 499 570 MBD.
 FT DNA_BIND 602 614 A.T hook 1.
 FT DNA_BIND 623 635 A.T hook 2.
 FT DOMAIN 800 865 DDT.
 FT DNA_BIND 1137 1149 A.T hook 3.
 FT DNA_BIND 1351 1363 A.T hook 4.
 FT ZN_FING 1623 1673 PHD-type.
 FT DOMAIN 1755 1825 Bromodomain.
 FT DOMAIN 613 738 Lys-rich.
 FT DOMAIN 647 774 Coiled coil (Potential).
 FT DOMAIN 1006 1012 Poly-Glu.
 SQ SEQUENCE 1850 AA; 205597 MW; E5DD4FED1D46DAE1 CRC64;

Query Match 24.08; Score 2437; DB 1; Length 1850;

Best Local Similarity 31.9%; Pred. No. 4.7e-84;

Matches 629; Conservative 309; Mismatches 573; Indels 461; Gaps 59;

Db 1174 CAVRAPITPGSYLYLRNTDTTKQKEP-----SDQLL-----NRWFSIVDK----- 1213
 QY 1378 THADMSTASLVTPQSQPPSPSPPTAPLGS---SAQNPVGNPLFALSPLQYKGVSMGM 1434
 Db 1214 -----ELPLSSTECPLPTINGSTPASLARQIFTNITCREICQIQGN----- 1254
 QY 1435 LQFCGWPVTGVTNIPFTLSVP-----SLGSLGLSEGNHSLFNTSNVASSKS 1482
 Db 1255 -----RWDIG---NNIQF-FSVPLEKGVETHFNEISILNSGSLDDDEINEVI-----AKK 1300
 QY 1483 ESPVP-----QNEKATGAQPAVAEVKXVPDF----- 1508
 Db 1301 SRPEPLQLSDMKPAFKRETIYSYQHPNQIRLRAEMWAETADEYGNFSLPAYWTL 1360
 QY 1509 -----PSPKPIPEEMQFGWRIIDPED 1530
 Db 1361 TLSNLTAYVQCDQFQPLQMTPEEKQLEDVKQHGAPQKTEPQVVPREFRYGWWKINDIEE 1420
 QY 1531 LKALKVLHLGIREKALQKQIKHLDYITQACLKNQDVAILNENENQVTRDIV--- 1587
 Db 1421 LNELIKALNPRGVRRLRQSL---LESIAES-----VNLTTPHVSHPPRAAPP 1466
 QY 1588 -----ENWSV---EQAMWMDLSVLQOVDELERVASLQVKGWMCPEPASE----- 1632
 Db 1467 PNGYIEPANWAFNSIARRVEVALDQIEAMEDKVASMQVQKQMPQREGDSENGV 1526
 QY 1633 -----REDLVYFEHKSFTKLCEHDEGFEFTGEDESSAHLERKSNPLDIAVTRLADL 1684
 Db 1527 EDVTIELRLERILGLEAATERRYLKPLGINTTEAQMAVIAQOESHQNQN---VSNLSNC 1584
 QY 1685 ENRIERTIEDIAPGLRWRRLSARSAAQVACICIQLOKSIWEKIMKYVCICIRKG 1744
 Db 1585 SNSSAE---DENLPKGLLSWRDAVERSVTTAQLSMALYVLESVAMDKSIMKANCFQCSG 1642
 QY 1745 DNEELLCLDCGCKCHYCHRPKLTTPDGDWFCPACIAKASGOTLKI----- 1793
 Db 1643 ESEDKLLLCDCGDRGHTYCFAPRMDKIPDGDWYCFECKNKATDRKICVCGLRPPPLG 1702
 QY 1794 -----KCLHVKGKK--TNESKK-- 1808
 Db 1703 KMWYCELCPRAYHQCYPPLMKYPRGKWKYQNCVAKAPPKKPQKPKERTTNNSSQSL 1762
 QY 1809 ----- 1808
 Db 1763 LNSSLNSSQNSLNSSHEDIATPLSATSAATVEQAOQSDYVSGEAAQYASGEYTHAQ 1822
 QY 1809 ----- 1808
 Db 1823 STSECEPQQSAPAEVDQIEASESPVPTAGGFEYFSPSNANNAAGSSTYSLDSDRTP 1882
 QY 1809 ---GKKVTLTGDTDE-----DS-----ASTSSSLKRGKNDLOKRKWEENTSINL 1850
 Db 1883 GGQDGEESCGDSSEYQPRPSPVDSLCLAGSSSSSSISDHOEKERSKERDEAKER 1942
 QY 1851 SKQSFSTSVKKPKRDSKDLALCSMLFEMTHEDAWFLPLVNLKLVPGYKVKIKKPM 1910
 Db 1943 AKQE-----KKATKRLLELAVCKTILEMELHEDSWPFLPVNTKQPTVYKVKIKSPMD 1997
 QY 1911 FSTIREKLSQYPNLETFALDVLVFNCFETFNEDSDDIGRACHNMRKYFEKKTWD 1967
 Db 1998 LSTIKRLQDLVYKSRDFIADVRQIFDNCVFVEDSDSPVGIAGHGMKFFEQRWAD 2054

RESULT 11
 Q9V602
 ID Q9V602 PRELIMINARY; PRT; 3080 AA.
 AC Q9V602;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE CG10897-PA.
 GN Name=tou; ORFNames=CG10897;
 OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 RN NCBI_TaxID=7227;
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
 RA Abrial J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Besley E.M.,
 RA Beeson K.V., Bencos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K.J., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrieres S., Fleischmann W.,
 RA Fowler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mallet B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J.R., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
 RA Shue B.C., Sriden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
 RA Williams S.M., Woodgett, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
 RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195 (2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426065; PubMed=12537568;
 RA Celniker S.E., Wheeler D.A., Kronmiller B., Carlson J.W., Halpern A.,
 RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,
 RA George R.A., Hoskins R.A., Lavery T., Muzny D.M., Nelson C.R.,
 RA Pacleb J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,
 RA Svirskas R., Taber P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
 RA Weinstock G., Scher S.E., Myers E.W., Gibbs R.A., Rubin G.M.;
 RT "Finishing a whole-genome shotgun: Release 3 of the Drosophila
 melanogaster euchromatin genome sequence.";
 RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426070; PubMed=12537573;
 RA Kaminker J.S., Bergman C.M., Kronmiller B., Carlson J., Svirskas R.,
 RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,
 RA Ashburner M., Celniker S.E.;
 RT "The transposable elements of the Drosophila melanogaster euchromatin:
 a genome perspective.";
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).
 RN [4]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426069; PubMed=12537572;
 RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
 RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,
 RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
 RA Bettencourt B.R., Celniker S.E., de Grey A.D., Drysdale R.A.,


```
QY 1189 RYMLPRCGGIFVEGMSG-----EGLEEI-----AKEREKKAESVOIK 1229
DB 1662 RYWLKPKAGGIFIEALSAQNDICDYHEALEAMDKDANDKENSENEKVAABESSQP 1721
QY 1230 BEMFET-----SGDSLNCSTHCEQEKLEKEKONTNL----- 1262
DB 1722 MEVDESITKLEGVASDVGPESNQNAHQDEEDDDDDVTINKVBEIVDLGDDDDA 1781
QY 1263 -----FLOKPGSFKSLKLEVAKMPP-----ESEVWTPK-----PNAGANGC 1300
DB 1782 APPLPKIEPPRPEIKVKSEMELMGPPPTMISTKTDFEAEIKIPSMGPGLMPPPTLNNNT 1841
QY 1301 TLSYONSGHS-----LGSVSTATOS-----NVEKADSNLFTNGSGGPKFYSP 1346
DB 1842 NNNNNNGNDCKLETGLGQQOQNFQSVIKTEKEDDCIIVTSS----- 1892
QY 1347 LPNDOLLTKLEKNQWFSLLPR-----TPCD-----DTSLTHADMSTASLVTPOS 1392
DB 1893 -----VDDTPKWSIVRREVLPISELPAEAGGVVGGQELQISYANQCSAQLOLQG 1942
QY 1393 QP-----PSKSPSPPTPAPLGSSAQNVPGLNPFALSPLOVQGGV----- 1430
DB 1943 HPWDLINNNQYVSIIPMDECKVDTSKLGNEC-----IFSLGLDEKQMLAKVBSYKAHK 1995
QY 1431 --SMGLQFCGWP-----TGVT-----SNIPF 1451
DB 1996 VESKNGL---GSPHRRHETKODEQAKLKDKEIDTEMETDADDLAGKEFFRLASDVP- 2051
QY 1452 TLSVPSLGSL-----GLSEGNSTSLTNSVASSKSSBP--VPQNEKAT 1493
DB 2052 -----PDTGGVSEGTDPVKIELRLDEALSAQYHNIANNLSVQTVIPIIDIPLEMT 2107
QY 1494 SAQPAAEVAKVPFPSPKP---IPEMQFGWRIIDPEDLKALKLVHLRGIREKALQK 1550
DB 2108 PDEHRLLEQVKLAGPFRVGVYPRQRQYGMWQLDDEQKLRLKTLNPSGLRERELEQE 2167
QY 1551 QIQXHLDIYTOACLKNKQVAIELNENENQVTRDIVENSWVEEQAMWMDLSVQOVEDL 1610
DB 2168 NLQRFGLQPLGVNYKLKSIDF---PEEFLLPDKGDMN-PKVAKRVELALIEQLBSL 2223
QY 1611 ERRVASASLOVKGWMCPE-----EPASE-----REDLVYFEHKSFTKLC 1648
DB 2224 EDKVASASMLKNQWLFNRVSELTLDQEDVTEEDFVSIIPMIRERIIDLEANIERYL 2283
QY 1649 KEHDEFTG-----EDESAAHA----- 1665
DB 2284 KPPLGSQTDLAHIAVIAQNHQTTTQTONSASAAVLLQWQQQQQQOQLAQOQQQQSG 2343
QY 1666 -----LERKSDNPLDIATVRLA----- 1682
DB 2344 AGNSLNPSSFNERTMALAAAAAASGPGNATGVANSVAVAGATPCESGSGEPNSGNASPAS 2403
QY 1683 DLERNIERRIEDTAPGLRVHRRALSEARSAAQVLCIQLOKSTANEKSTMKVY----- 1737
DB 2404 NCDSDRDEKVEQ-IPKGLVQMRDAVSRSHHTAQALMALYLVLESVCAWDKSIKAYATQNK 2462
QY 1738 ----- 1737
DB 2463 SSKKSSAKQATPSKKQQQKNKKEQQLTPNGKESKSAINKPENQAPLSIKINKAL 2522
QY 1738 ----- 1737
DB 2523 AQNGOQLKTKTPPILTKPTASPSHSPNTSDSDFGKRTKKSGKRRRSANNNTSSK 2582
QY 1738 -----COICRKGDNELLALCGCDKGYCHRPKITTIIPGDWFCPACIAKASQ- 1789
DB 2583 YSNLQNCQFTSGENEDKLLLCDCGDKGYHTYCFKPKQMDNIPGDWYCYECVNKATNER 2642
QY 1790 -----TLKI-----KKLHV 1798
DB 2643 KCIVCGGHRPSVGMIMYCDLCPRAYHADCVIPELLKVPGRKWTCHGCISRAPPKKRS 2702
QY 1799 KGKKTNESKKGKVTLTGDTDEDEDSASTSSSLKRGN-----KDLQKRKN----- 1842
```

```
DB 2703 GGTSGSSKSR-----DRRESGSAKRRSDNSKTPAMEHMQQQOQPLAGDSSH 2753
QY 1843 ---EENTISINLSKQBSFTS----- 1858
DB 2754 HHHQQPPPLNSHSDSMMSLPAPLSPAHSVVSATNYDDOHHANNNSVDGSSRFHAHLIP 2813
QY 1859 ----- 1858
DB 2814 SNNGTAALLEDPVGGANVMPGVVYVYTPVAAGNFSAGLINQAPVQAMPFANVAMSPRA 2873
QY 1859 ----- 1858
DB 2874 VTPTRTRTPPTPAPTPPPPTPLMQASPTATALHVNACOSPQQOQHAQLMTWPPSPA 2933
QY 1859 -----VKPKRDD-----SKDLALCSMLTETHEDA 1886
DB 2934 IGVGTATNQMSPPPINIHAIQAEKLEKQEKKEKATKKMLKELAVCKTLGEMBLHEDS 2993
QY 1887 WPFLLPVNLKLVPGYKVKIKKPMDFSTIREKLSGQYPNLETFALDVLRFVONCETFNED 1946
DB 2994 WPFLLPVNTKQPPTYRKIIKTMDLSTIKKQLDLSYKTREDFCVDVQRFQNCMFNED 3053
QY 1947 DSDIGRACHNRKYEKKW 1965
DB 3054 DSPVGKAGHGRKPFESRW 3072

RESULT 12
Q9BMQ0 PRELIMINARY; PRT; 3109 AA.
AC Q9BMQ0;
DT 01-JUN-2001 (TREMELrel. 17, Created)
DT 01-JUN-2001 (TREMELrel. 17, Last sequence update)
DT 01-MAR-2004 (TREMELrel. 26, Last annotation update)
DE Toutatis.
GN Name=tou;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_taxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21589546; PubMed=11732091;
RA Fauvarque M.O., Laurenti P., Boivin A., Bloyer S., Griffin-Shea R.,
RA Bourbon H.M., Dura J.M.;
RT "Dominant modifiers of the polyhomeotic extra-sex-combs phenotype
RL induced by marked P element insertional mutagenesis in Drosophila.";
RL Genet. Res. 78:137-148(2001).
RL EMBL; AF314193; AAK00302.1; -.
DR HSSP; Q9UIG0; 1f62.
DR FlyBase; FBgn0033636; tou.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR000194; ATPase a/bcentre.
DR InterPro; IPR001487; Bromodomain.
DR InterPro; IPR004022; DDT.
DR InterPro; IPR001739; Methyl-CpG_bind.
DR InterPro; IPR006162; Ppantne_S.
DR InterPro; IPR001965; Znf_PHD.
DR InterPro; IPR001876; Znf_RangDP.
DR Pfam; PF00439; Bromodomain; 1.
DR Pfam; PF02791; DDT; 1.
DR Pfam; PF01425; MBD; 1.
DR Pfam; PF00628; PHD; 2.
DR PRINTS; PR00503; BROMODOMAIN.
DR SMART; SM00297; BROMO; 1.
DR SMART; SM00547; Znf_RBZ; 1.
DR PROSITE; PS00152; ATPASE_ALPHA_BETA; UNKNOWN_1.
DR PROSITE; PS50014; BROMODOMAIN_2; 1.
DR PROSITE; PS50827; DDT; 1.
```



```
DR PROSITE; PS00012; PHOSPHOPANTHETINE; UNKNOWN_1.
DR PROSITE; PS01359; ZF_PHD_1; UNKNOWN_1.
DR PROSITE; PS00016; ZF_PHD_2; 2.
SQ SEQUENCE 3109 AA; 33242 MW; DD0A75215541DCED CRC64;

Query Match 14.3%; Score 1452; DB 2; Length 3109;
Best Local Similarity 20.2%; Pred. No. 1.5e-46;
Matches 605; Conservative 382; Mismatches 753; Indels 1248; Gaps 82;

Qy 2 GQTSTSSGGNRKCNQEQSNQPLDARVDKIKDKPKKAKMSSSSSDSG-----T 55
Dy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dy 338 GSGGKSGGGGS-----SSSMHSLGMGIGSGGGMSSGKASTSVASNSLGGMHPSLT 392
Qy 56 SSD---TSSEGISSSDDLEEDDEEDQSEEDDDDDSEEAQKNNOVLLHGISD 112
Dy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dy 393 SSNPLMSPHAGMGSSS-----SSSKDRDKNNPSLNALNSLSQFALGWT- 437
Qy 113 PKADQKATEKAQEKRIHQPLAPESQTHSFQSQ----- 148
Dy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dy 438 PQSQMAAMNAPAAASTGVSPSATVTSPPHSSQQQQQMGNSSTSGSGKSSKDYMMGTG 497
Qy 149 -KQPVLSQQLPPIFQSQAKESVNHKTSVIQSTGLVSNVKPLSLVNOAKKETYMKLIV 207
Dy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dy 498 SEHPSLGLVRLP-----PDTEIITYTSIVGPKI----- 526
Qy 208 PSPDVLKAGNKNTSESSLLTSELRSKREYKQAPPSQOLKKOESSK---SLKKVIAA-- 261
Dy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dy 527 --PGTTSRGRKKTISETEQTTQQQK--QQHQAQHLQQQQQAQKELDTKNAISLL 582
Qy 262 ----LSNPK-----ATSSSPAHPKQTLNHNPNPFLTNALLGNHQPNGVTSV--- 305
Dy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dy 583 APFGLSPAKRLARLEWYAAMAAQAQQHQQAQQQAQQHQAQQHQAAGIPGMAGLVGLP 642
Qy 306 -IQAPALATTKTKQKQKINENIAAASSTPPSPVNLSTSGRRTPGQNTVPWPSAPI-L 363
Dy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dy 643 GMSGNPLDQLSVSKASSSTAPTTSSTSSAGSNLLNQNSDRVEVILPPTITSNGAYNL 702
Qy 364 HSQCKEKAVSNVNPVKTQ--HHHPAKSLVEQFRGTDSDIPSSKSDSEDSNEDEED-DE 420
Dy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dy 703 SSKGK-----VHDLTTDMATNGGVNLSKSNAGSALTSPGAVGASAPITIDPDA 756
Qy 421 EDEDEDEDDSDSQSDESNSSDTEGSEEDDDKQDQSDSDTEGKTSMLNKTT 480
Dy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dy 757 PLNLSMKPSDKSNSSSNAAGGSSSSAALANLADSY---QAASSGSGSNLSQLSLIT 813
Qy 481 SSVKS-----PMSLTG-----HSTPNL--HI 501
Dy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dy 814 AALGGTGMPPGGSISGSGGTSPAPAGAGSGATGGSGSGSGSSSYKEGRPNLGRGV 873
Qy 502 AKAPGSAPAALCSE-----SOSPAFLTSSSTLTSS 532
Dy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dy 874 SKPKNTVSLAQSRAVGLKPMLATQQLQOGADIKIRLALSEANAHTSTDSSEVA 933
Qy 533 PHSGTSKRR-----RVTDERELRIPLYGWQRETRINF--GRLQGEVAYYAP-CGKKLR 585
Dy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dy 934 AESGLSESESDANILNVAELRVPLELGNKRETVIRGLTKQGIQIRGEVYIYAPGSTTPLK 993
Qy 586 QYPEVIKYLNRNGIMDTSRDNPFSAKIRVGDFTVEARDGPOEMQCLLKEDDVIPRIAM 645
Dy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dy 994 SNGQVFAILLEQOP--SNLSRENFSFARAIVGSFLQ----- 1027
Qy 646 EGRGRPNPDRQAREESRRRRKGRPNPVNGNAEFLDNADAKLLRKLQAOE1-ARQAAQ 704
Dy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dy 1028 -----PAP-----PPYANDGEYIRMTDEDVAKRLUEDLKVTFTRQLN 1063
Qy 705 IKLLRKLQKQBARVAKEAKQQA1MAAEKRRKQEQIKIMKQEQEKRIQOIRMEKELR 764
Dy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dy 1064 V-----EQRLEIAKQ-----QAMRDAKKLQ 1084
Qy 765 AQQILEAKKKKEAANNAKLEAEKRIKEKEMRQQOAVLLKHQERRRRQHMLMKMEA 824
Dy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dy 1085 KEELARNEKARQE-KNSKL---EQQRDKELKNQQAQV-----EBERRRRQHSIIRMLEL 1136
Qy 825 RKKAEXERLUKQEKREKRLNKKERKLBQRLELEMAKELKKPNEDMCLADQKPLPELPI 884
Dy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dy 1137 RKPEDEKKGHLQVLDRLLLRERRMAERKADABILQIRRNEDSEMPQLVPELDRI 1196
Qy 885 PGLVLSGSTSFSDCLMVQFLNFKGVLFQVNDVFNLSVLQEGILL--NIGDSMGVEQDL 942
Dy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dy 1197 AGNELPGQAMADLLMVFEFLHNFGETLGFOME-SLPSLQNLHLDALMSDNADABEELSV 1255
Qy 943 LVRLLSAAVCDPGLITGYKAKTALGHELLNVGNVNRDNVSEILQIFMEA----- 990
Dy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dy 1256 MTHLLVCAIEDPGVPNFGRRHTLLGQSLRNADITNSVNSEILRIYLYATATGEVRQMGI 1315
Qy 991 -----HCGQT-ELTESLTKAKFAQHTPAQ 1013
Dy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dy 1316 TVDRERRRVDHHQLSDTTTHSHSVKNQYVYKLLHENDTWKLSQSLKDRPFFVALNPT 1375
Qy 1014 KASVLAFLINELACSKSVSEIDKNIDYMSNLRDQKVVVEGKLRKLRIRIHAHK----- 1066
Dy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dy 1376 KAOMLAHLCDLLMNAVLRLQIDGSLCTCAQMRKEKYMMDMKVRKYKALHMRKARIEAYE 1435
Qy 1067 -----TKCRDTSGLDGE-- 1080
Dy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dy 1436 RAQAREAMQALMAOQKLDARLUKABEEAKAAAAAAGTDGEATKGGSPNGEAP 1495
Qy 1081 -----EQHPL----- 1085
Dy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dy 1496 DGDQNEGAKEPQQQQQPMEDVGVVDEASLVSPAKTIIQTDNSLTPSKQDMPTPTYQI 1555
Qy 1086 -GTPTPGR-----KRRRKGDSYDDDDDD-----SDQGDDEDEDEEDKEDQ 1128
Dy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dy 1556 NGSTPTTSGVTGDMNVLLQAKKSGARNSINDEHHHDVSIIDDDLSLDSE----- 1607
Qy 1129 KGKKTDCEDDEGDQAASVEELEKQIEKLSKQSQVRRKLFDAHSLSRVMGFDPYRR 1188
Dy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dy 1608 ---ITNVEED---DNRLSADELOKLDKIVRASLNCKEALKEKSTNQURAAACFGQDRFWR 1661
Qy 1189 RYWTLPRCGGIFVFGMESG-----EGLPEI-----AKEREKLUKAEVSQIK 1229
Dy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dy 1662 RYWKLPKAGGIFIEALESQNDICDYHEALEAMDKKDANDEKENSENEKDVAARESSQP 1721
Qy 1230 EEMFET-----SGDSLNCNTDHCQEKDELKEKONTNL----- 1262
Dy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dy 1722 MEVDSEITKLEDGVPASDVGMPSNQONAHQDEDDDDDDVTEINKRVEPEIVDLGDDDDA 1781
Qy 1263 -----FLQKGSFSLKSLLEAVKWP-----ESEVWTPK-----PNAGANGC 1300
Dy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dy 1782 APPLPKTEPPRPEIKVSEMELMGPPPTMTISTKTDFEAEIKI FSPMGLMPPPTLNNT 1841
Qy 1301 TLSYQNSGKHS-----LGSVQSTATQS-----NVEKADSNLNFNTSGSGPKFYSP 1346
Dy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dy 1842 NNNNNNGSDCKLETGLGLGQQQNFQSVIKTEDVKKEDDCIIVSTSS----- 1892
Qy 1347 LPNDQLLTKLRQWPSLLPR-----TPCD-----DTSLTHADMSTASLVTQS 1392
Dy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dy 1893 -----VDDTPKWFIVRREVPLISELPABEGGVQGELOQTSYANQCSAQLQGG 1942
Qy 1393 QP-----PSKSPSTPAPLGSSAQNPVGLNPFLSPQVKGCV----- 1430
Dy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dy 1943 HPWDLINMYYISPMDECKVDTSKLGNEC-----IFSLGDEKQMLAKVEYKAHK 1995
Qy 1431 --SMWGLQFCWGP-----TGWVT-----SNIPF 1451
Dy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dy 1996 VESKNGL---GSPRRHETKDDDEQAKLKDKEIDTEMETDADDLAGKEKFRLSRDVP- 2051
Qy 1452 TLSVPSLSGSL-----GLSGNGNSFLTSTNVASSKSESP---VPONEKAT 1493
Dy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dy 2052 ---PDTGGGVSEGTVDVKPIELRLDEALSQAVYHNIANMSSLSSVQTYIPIDPLSMT 2107
Qy 1494 SAQPAAEVAKVPDPSPKP---IPEEMQFCGWRIIDPEDIKALLKVLHLRGIREKALQK 1550
Dy : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Dy 2108 PDEHRLLEQVKLAGFFPERVHGVIYVRRQRYGWWQDDDEQKLRQLLKTUNPSGLREREQ 2167
Qy 1551 QIQKHLDYITQACLKNKDVAIIELNENENQVTRDIVENWVSEEQAMEMDLSVLOQVEDL 1610
```



```
Db 2168 NLQRFGLGLEQPGVNYKUKSDIDF---PEEFMPDKKGDWN-PKVAKVVELALIEQLRESL 2223
Qy 1611 ERRVASASLQVGMWCP-----EPASE-----REDLVYFEHKSFTKLC 1648
Db 2224 EDKVASASMLQVWQPNRVESELTDQSDVTEEDFVSIIPMIRERIIDLEANNERYL 2283
Qy 1649 KEHDEFTG-----BDESSAH-----1665
Db 2284 KPPLGSQTGDARLAVIAQNHQHTTTOTONSASAAAYLLQWQOQOQOQOQOQOQOQSG 2343
Qy 1666 ----LERKSDNPLDIATVRLA-----1682
Db 2344 AGNSLNPSFPIERTWALAAAAAASGPNATGVANSVAVAGATPCSGSGGPNVGNASPAS 2403
Qy 1693 DLERNIERREDIAPGLRVRRALSEARSAAQVALCIQLOKSTAWKSIKMYV-----1737
Db 2404 NCDSRDEKVEQ-IPKGLVQWRDAVSRSHTTAQLAMALVYLVESCVADKSIKAVATQK 2462
Qy 1738 -----1737
Db 2463 SSKKSSAKQATPSKQOQKNKKEQQLTPNGKESKSAINKPENQAPLSIKINKAL 2522
Qy 1738 -----1737
Db 2523 AQNGQLKTKTPPILTKSPASPSHPNTSDSDSDFGKRTKKSGGKRRRSANNTNSSK 2582
Qy 1738 ----CQICRKGNEELLCDGCKGCHTYCHRPKITTPDGDWFCPACIAKASQ- 1789
Db 2593 YSNSLQNCQFCTSGENEDKLLCDGCKGYHYCYFKPMDNIPDGDWYCYECVAKATNER 2642
Qy 1790 -----TLKI-----KLLHV 1798
Db 2643 KCIVCGHRPSFVGKMYCDLCPRAYHADCIYIPLLKVPGRGWKYCHGCSIRAPPPKRS 2702
Qy 1799 KGKTNESKKGKVTLTGDTDEDSASTSSLSKGN-----KDLQKRM-----1842
Db 2703 GGTSGSSSKSR-----DRRESGSGAKRRSDNSKTPAMEHMQOQOMPLAGDSSH 2753
Qy 1843 ---BENTSINLSKOBESFTS-----1858
Db 2754 HHQQPFLNSHDESMNSLPAGSSEVSAGCRLNGRRSTRSGFQLNNSALCSPAHSV 2813
Qy 1859 -----1858
Db 2814 VSATNYDDQHANNVSDGSSRFHAHLIPPSNNGTAALEDVPGGANVMPGVVPTVPA 2873
Qy 1859 -----1858
Db 2874 GNFSAGLINQAPVQAMPFANVAMS PRAVTPTRTRTPPTPAPTPPPPPTPLMQASP 2933
Qy 1859 -----VKKPK 1863
Db 2934 TATALHVNACOSPPQOQAQLMTWPSPPAIGVGTATNQMSPPPIINIAHQBAKELKQEK 2993
Qy 1864 RDD-----SKDALCSMLTEMETHEDAWFLFLPVNLKLVPGYKVKIKKEMDFSTIREK 1917
Db 2994 KEKHATKLMKELAVCKTLGEMELHEDSWFLLPVNTKQFPPTYRKIIKTMDLSTIKK 3053
Qy 1918 LSSQYQPNLETFALDVLRFONCFETNEDDDSDIGRAGHNMRYFEKKW 1965
Db 3054 LQDLSYKTRDFCDVVRQIFQNCMFENEDDPSVGKAGHMKRFFESRW 3101

RESULT 13
Q80U42
ID Q80U42 PRELIMINARY; PRT; 886 AA.
AC Q80U42;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE MKIAA0314 protein (Fragment).
GN Name=Baz2a; Synonym=mkIAA0314;
```

```
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=22579291; PubMed=12693553;
RA Okazaki N., Kikuno R., Ohara R., Inamoto S., Aizawa H., Yuasa S.,
RA Nakajima D., Nagase T., Ohara O., Koga H.;
RT "Prediction of the coding sequences of mouse homologues of KIAA gene:
RT II. The complete nucleotide sequences of 400 mouse KIAA-homologous
RT cDNAs identified by screening of terminal sequences of cDNA clones
RT randomly sampled from size-fractionated libraries.";
RL DNA Res. 10:35-48(2003).
DR EMBL; AK122243; BAC65525.1; -.
DR HSSP; Q9UIG0; IFG2.
DR MGD; MGI:2151152; Baz2a.
DR GO; GO:0006338; P:chromatin remodeling; TAS.
DR InterPro; IPR000637; A+T_hook.
DR InterPro; IPR001487; Bromodomain.
DR InterPro; IPR011011; FYVE_PHD_ZnF.
DR InterPro; IPR001965; ZnF_PHD.
DR Pfam; PF02178; AT_hook; 2.
DR Pfam; PF00439; Bromodomain; 1.
DR Pfam; PF00628; PHD; 1.
DR PRINTS; PR00503; BROMODOMAIN.
DR SMART; SM00384; AT_hook; 2.
DR SMART; SM00297; BROMO; 1.
DR SMART; SM00249; PHD; 1.
DR PROSITE; PS50014; BROMODOMAIN_2; 1.
DR PROSITE; PS50016; ZF_PHD_2; 1.
FT NON_TER 1
SQ SEQUENCE 886 AA; 100060 MW; 43E1172F5F081E06 CRC64;

Query Match 13.1%; Score 1331; DB 2; Length 886;
Best Local Similarity 34.0%; Pred. No. 1.4e-42;
Matches 346; Conservative 137; Mismatches 295; Indels 240; Gaps 37;

Qy 1055 KLRKRIIHAKTGKTRDT--SGGIDLGEQHPGLTPTGKRRRRKGGSDYDDDDDDSD 1112
Db 1 KLRKRLKALAKTGRPEVMGEAD-----GLGRRSSRIME-----E 38
Qy 1113 DQGDDEDEDEBK---EDQKGKTKTDICEDEGDQAA-SVEELEKQIEKLSKQSQYRRK 1168
Db 39 TSGIEEEEEEENTTAVHGRGRK-----EGEIDVAASSIPELERHIEKLSKRLQFFRK 92
Qy 1169 LFDASHLSRVMPGPDYRRRYRWILPRCGGIFVGMESGEGLE-EIAKERELKK-----1222
Db 93 LLHSSQMLRAVSLGQDRYRRHYVLPYLAGIFVSGSEGSTVTEDEIKQETESLMEVVTST 152
Qy 1223 --AESVQIKEMPFETSGDSLNCSTNDHCEQKEDLKERDNTNLFLOKPGSPSKL-----1273
Db 153 PSSARASVKREL-----TGSNASTSPARSGRPRK-----PKGSLQFQHLQSTIR 198
Qy 1274 ---SKLLEVAKMP-PESEVWTP-KPNAGANCTLSYQNGKHSGLSGVSTATQSNV-----1324
Db 199 ECDSEQAQTQVHPFQQLQAPTOHLPQSPSGFLEPEGS-PFSLGQSQHDLSQSASFLSWL 257
Qy 1325 -EKADSNLNFN-----TGSSSGPKFYS-----PLPNDQLLKLTKRNQWFSLLPRTP 1371
Db 258 SQTQSHNSLLSSSVLTPDSSPGKLDASPSQLESPEPDE--AQSCPGQPGWFNPSAQIP 315
Qy 1372 CDDTSLTHADMSTASLVTPOSQPPSKSPSTPA-----PLGS-SAQNVPGLNPFALS 1422
Db 316 CD-----AAPTTPPAVSDQPTPFSQLLASSKPMNTPGAANP--CSPVQLS 359
Qy 1423 PLQVKGVSVMGLQFCGWPTGVVTSNTPFTLSVPSLGLSGLSEGNNSFLTNNVASSKS 1482
Db 360 STLHPGGT-----PKRLS-----GDS-----EEMS 379
Qy 1483 ESPVPQNEKATSAQPAAVEVAKPVDFP-----SPKPIPEEMQFGWRIIDPEDLKALKV 1537
Db 1483 ESPVPQNEKATSAQPAAVEVAKPVDFP-----SPKPIPEEMQFGWRIIDPEDLKALKV 1537
```

```
Db 380 QSPITGLQPKRRGRPPS-KFFKQVEQHYLTQLTQTAQIPPEMCSGWMWIRDPTETLDVLLKA 438
Qy 1538 LHLGIRKALQKIQKLDYITQACLNKNDVAIELNE---NEENQVTRDIVENASVEE 1594
Db 439 LHPGIRKALHKLHSLRKHQFLQEVCLQPLTDPFEPNELPALBEG-----VMSWSPE 492
Qy 1595 QAMEMDLSVLOQVEDLERRVASASLQVGMWCPBPASERDLVYFEHKSFTKLCKKEHDE 1654
Db 493 KTYETDLAVLQWVELEQORVVLQDQIRGWTCTPFDSTREDLTYCEHLFDS----PEDIP 548
Qy 1655 FTGDESSAHALERKSNPLDIATVRLADLERNIERR----- 1691
Db 549 WRGREGRET-VPQRQNNPLDLAVRLAVLEQNVRRYLREPLWAAHEVVVEKALLSTPN 607
Qy 1692 -----IIEDIAPGLRVRRALSSEARSAAQVLCIOLOKSIANEKSIIMKYQCICR 1742
Db 608 GAPDGTSTEISYEITPRVVRWQRTLCRSAAQVCLCGQLERSIAWEKSNKYTCLVCR 667
Qy 1743 KGDNEELLLCDGCDKGCHTYCHRPKITTIPDGDFWFCPACIAKASGOTLKIKLHVKGKK 1802
Db 668 KGDNEFLLCDGCDRGCHYCHRPKMEAVPEGDFWFCVCLSVQVEEYTORPGPKRGQK 727
Qy 1803 TNESKKGKXVLT---GDTEDDSASTSSSLKRGNKDQKRKM-----EENTSINLSKQES 1855
Db 728 RKSS-----FPLTPPEGDS-----RRMLSRGRDSPAVPYPEDG 762
Qy 1856 FTSVKKPK---RDSKDLALCSMLTSMETHEDAWPELLPNLKLVPQYKVKVKKPMDFS 1912
Db 763 LSPPKRRHRSRHHSDLTFCIILMEMESHDAWPFLEPNPRLVSGYRVKVIKNPMDFS 822
Qy 1913 TIREKLSGGQPNLETFALDVLRFVDCETFNEDSDIGRAGHNRKYFEKKWTDTPK 1970
Db 823 TMRRELLRGYTSSEFAADALLVFDNCQTFNEDDSEVGKAGHVRFFESRWEFFQ 880

RESULT 14
Q6MZK7
ID Q6MZK7 PRELIMINARY; PRT; 441 AA.
AC Q6MZK7
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DE Hypothetical protein DKFZp686H10114.
GN Name=DKFZp686H10114;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Human fetal kidney;
RG The German Human cDNA Consortium;
RA Koehrer K., Beyer A., Mewes H.W., Weil B., Amid C., Osanger A.,
RA Fobo G., Han M., Wiemann S.;
RL Submitted (AUG-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL: BX641037; CAB46023.1; -.
KW Hypothetical protein.
SQ SEQUENCE 441 AA; 46936 MW; 08ECC8A9EA7772B9 CRC64;

Query Match 11.5%; Score 1169; DB 2; Length 441;
Best Local Similarity 98.3%; Pred. No. 8.3e-37;
Matches 235; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 MGQTKSTSSGGNPKNCQEQSKNOPLDARVDKIKDKPKRKAMESSNSDSGTSDDTS 60
Db 195 MGQTKSTSSGGNPKNCQEQSKNOPLDARVDKIKDKPKRKAMESSNSDSGTSDDTS 254
Qy 61 SEGISSDSDLDLEDEEEQDSIESDDSDSESAQHKSNNQVLLHGISDPKADGOKA 120
Db 255 SEGISSDSDLDLEDEEEQDSIESDDSDSESAQHKSNNQVLLHGISDPKADGOKA 314
Qy 121 TEKAQEKRIHOPLPASEQTHSFQSQKQPVLSQQLPFIFQSSQAKESVNVKHTSVIQ 180
```

```
Db 315 TEKAQEKRIHOPLPASEQTHSFQSQKQPVLSQQLPFIFQSSQAKESVNVKHTSVIQ 374
Qy 181 STGLSVNVKPLSLVNOAKKETMYKLIYPSDPVLKAGNKVTSSESSLITSELRSKREQYK 239
Db 375 STGLSVNVKPLSLVNOAKKETMYKLIYPSDPVLKAGNKVTSSESSLITSELRSKRVSYK 433

RESULT 15
Q23590
ID Q23590 PRELIMINARY; PRT; 1376 AA.
AC Q23590;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Flectin protein 1.
GN Name=Flt-1; ORFNames=ZK783.4;
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RG MEDLINE=99069613; PubMed=9851916;
RG WormBase Consortium;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Favello A., Vaudin M.;
RT "The sequence of C. elegans cosmid ZK783.";
RL Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Waterston R.;
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Wilson R.;
RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Wilson R.;
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RG WormBase Consortium;
RL Submitted (SEP-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL: U13646; AAC24421.2; -.
DR PIR: T34516; T34516.
DR HSP; Q14839; IWM3.
DR WormBase; WBGene00001470; flt-1.
DR WormPep; ZK783.4; CE34152.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR001487; Bromodomain.
DR InterPro; IPR004022; DDT.
DR InterPro; IPR001739; Methyl-CpG_bind.
DR Pfam; PF00439; Bromodomain; 1.
DR Pfam; PF02791; DDT; 1.
DR Pfam; PF01429; MBD; 1.
DR Pfam; PF00628; PHD; 2.
DR PRINTS; PRO0503; Bromodomain.
DR SMART; SM00297; BROMO; 1.
DR SMART; SM00571; DDT; 1.
DR SMART; SM00391; MBD; 1.
```


AC Q9NRL2; Q9NZ15; Q9P065; Q9UIG1; Q9Y3V3;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Bromodomain adjacent to zinc finger domain protein 1A (ATP-utilizing
DE chromatin assembly and remodeling factor 1) (hACF1) (ATP-dependent
DE chromatin remodeling protein) (Williams syndrome transcription
DE factor-related chromatin remodeling factor 180) (WCRT180) (hWALp1)
DE (CHRAC subunit ACF1) (HSPC317).
DE Name=BAZ1A; Synonyms=ACF1, WCRT180;
GN Homo sapiens (Human).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1), AND MASS SPECTROMETRY.
RC TISSUE=Cervical carcinoma;
RX MEDLINE=20341093; PubMed=10680450; DOI=10.1093/emboj/19.13.3377;
RA Poot R.A., Deliaire G., Huelsmann B.B., Grimaldi M.A., Corona D.F.V.,
RA Becker P.B., Bickmore W.A., Varga-Weisz P.D.,
RT "HuCHRAC, a human ISWI chromatin remodelling complex contains hACF1
RT and two novel histone-fold proteins.";
RL EMBO J. 19:3377-3387(2000).
[2]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RX MEDLINE=20122567; PubMed=10655480; DOI=10.1073/pnas.97.3.1038;
RA Bochar D.A., Savard J., Wang W., Lafleur D.W., Moore P., Cote J.,
RA Shiekhattar R.;
RT "A family of chromatin remodeling factors related to Williams syndrome
RT transcription factor.";
RL Proc. Natl. Acad. Sci. U.S.A. 97:1038-1043(2000).
[3]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Testis;
RX MEDLINE=20101112; PubMed=10662543; DOI=10.1006/geno.1999.6071;
RA Jones M.H., Hamana N., Nezu J.-I., Shimane M.;
RT "A novel family of bromodomain genes.";
RL Genomics 63:40-45(2000).
[4]
RP SEQUENCE OF 1-541 FROM N.A. (ISOFORM 1).
RC TISSUE=Blood;
RA Ye M., Zhang Q.H., Zhou J., Shen Y., Wu X.Y., Guan Z.Q., Wang L.,
RA Fan H.Y., Mao Y.F., Dai M., Huang Q.H., Chen S.J., Chen Z.;
RT "Human partial CDS from cd34+ stem cells.";
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
[5]
RP SEQUENCE OF 746-1556 FROM N.A. (ISOFORM 1).
RC TISSUE=Uterus;
RA Wambutt R., Heubner D., Mewes H.-W., Gaassenhuber J., Wiemann S.;
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: May play a role in transcriptional regulation. May be
CC involved in the formation or maintenance of heterochromatin
CC playing a critical role in developmental control.
CC -!- SUBUNIT: Together with CHRAC1, POLE3 and ISWI/SNF2H proteins, it
CC forms the ISWI chromatin remodelling complex, CHRAC.
CC -!- SUBCELLULAR LOCATION: Nuclear. May target the CHRAC complex to
CC heterochromatin.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=1;
CC IsoId=Q9NRL2-1; Sequence=Displayed;
CC Name=2;
CC IsoId=Q9NRL2-2; Sequence=VSP_000551;
CC -!- TISSUE SPECIFICITY: Highly expressed in testis and at low or
CC undetectable levels in other tissues analyzed.
CC -!- MISCELLANEOUS: Stimulated by double-stranded DNA and nucleosomal
CC DNA.
CC -!- SIMILARITY: Belongs to the WAL family.
CC -!- SIMILARITY: Contains 1 bromodomain.
CC -!- SIMILARITY: Contains 1 DDT domain.
CC -!- SIMILARITY: Contains 1 PHD-type zinc finger.
CC -!- SIMILARITY: Contains 1 WAC domain.
CC -!- CAUTION: Ref.4 sequence differs from that shown due to multiple

frameshifts starting at position 382.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; AF213467; AAF70601.1; -
CC EMBL; AF221130; AAF32366.1; -
CC EMBL; AB032252; BAA89209.1; -
CC EMBL; AF161435; AAF28995.1; ALT_SEQ.
CC EMBL; AL050089; CAB43261.1; -
CC PIR; T08738; T08738.
CC HSP; Q9UIG0; IP62.
CC Genew; HGNC:960; BAZ1A.
CC MIM; 605680; -
CC InterPro; IPR001487; Bromodomain.
CC InterPro; IPR004022; DDT.
CC InterPro; IPR001965; Znf_PHD.
CC InterPro; IPR001841; Znf_ring.
CC Pfam; PF00439; Bromodomain; 1.
CC Pfam; PF02791; DDT; 1.
CC Pfam; PF00628; PHD; 1.
CC PRINTS; PR00503; BROMODOMAIN.
CC SMART; SM00297; BROMO; 1.
CC SMART; SM00571; DDT; 1.
CC SMART; SM00249; PHD; 1.
CC SMART; SM00184; RING; 1.
CC PROSITE; PS00633; BROMODOMAIN_1; 1.
CC PROSITE; PS00634; BROMODOMAIN_2; 1.
CC PROSITE; PS00827; DDT; 1.
CC PROSITE; PS01359; ZF_PHD_1; 1.
CC PROSITE; PS00016; ZF_PHD_2; 1.
CC Alternative splicing; Bromodomain; Coiled coil; Nuclear protein;
CC Transcription regulation; Zinc-finger.
CC DOMAIN 22 128
CC DOMAIN 422 487
CC ZN_FING 1148 1198
CC DOMAIN 1446 1516
CC DOMAIN 306 397
CC DOMAIN 634 709
CC VARSPLIC 504 535
CC CONFLICT 135 135
CC CONFLICT 236 236
CC CONFLICT 551 551
CC CONFLICT 730 730
CC CONFLICT 769 769
CC CONFLICT 1201 1201
CC CONFLICT 1206 1206
CC CONFLICT 1366 1366
CC CONFLICT 1409 1409
CC SEQUENCE 1556 AA; 178674 MW; 43F2667C3CCFD458 CRC64;
R -> T (in Ref. 3).
D -> E (in Ref. 3).
D -> E (in Ref. 2 and 3).
V -> F (in Ref. 3).
P -> L (in Ref. 3).
S -> C (in Ref. 3).
S -> F (in Ref. 3).
K -> N (in Ref. 2 and 3).
K -> R (in Ref. 2 and 3).
/FtId=VSP_000551.
R -> T (in Ref. 3).
D -> E (in Ref. 3).
D -> E (in Ref. 2 and 3).
V -> F (in Ref. 3).
P -> L (in Ref. 3).
S -> C (in Ref. 3).
S -> F (in Ref. 3).
K -> N (in Ref. 2 and 3).
K -> R (in Ref. 2 and 3).
7.1%; Score 721.5; DB 1; Length 1556;
Best Local Similarity 21.1%; Pred. No. 2.9e-19;
Matches 340; Conservative 182; Mismatches 424; Indels 663; Gaps 53;
QY 664 SRVRRKGRPNVGNAEFLDNADAKLLRLKLOAEIARQAQIKLLRLKLOKQEQARVAKA 723
DB 270 SPANRRGRPP-----KRIHSQSDNVA--- 292
QY 724 KQQAATAAEKKEKQEQIKIMQKQEKIKRIQOIRMEKELRAQOILEAKKKKEEAAANAK 783
DB 293 --NKQTLASYSRKATKEDKLLKQEE-----MKSLAFE-----KAKLKREKADA- 334
QY 784 LLEAEKRIKEKEMRROQAVLLKHQERRRRQHMMLKAMKAEKAEKELKQEKREKR 843
DB 335 -LEAKK--KEKE-----DKEKKREELKIVEERLKKKBEKELKVEREKE-- 377

QY 844 LNKERKLEORLEMAKELKPNEDMCLADOKPIPELPRIGLVLSGTSFSDCLMVOP 903
Db 378 --REKLEEKRYKVEYLKQWSPKPRDEMBECDDIKELPE-PTPVKTRLPPEIFGDALMVLFE 434
QY 904 LRFKGLVDFVDNIDVPN---LSVLQEGLLNIGDSMGEVQDILLRLLSAV----- 951
Db 435 LNAFGLF---FDLQDFPDGVTLEVLLEALVG-NDSEGPLCELLFFFLTAIFQAIABEEREE 491
QY 952 -----COPGLITGYKAKTAL-----GEHLLNVGV 975
Db 492 VAKEQLTADTKDLTEALDEDADP-----TKSALSASVLAAMWOLHQCSLKSLDL 544
QY 976 NFDNVSEILQIF-----MEAHCGGTETLESKTKA 1005
Db 545 DCSLTSLDILRHILASGADVTSANAKRYQKRGFGFATDDACMELRLNPSLVKLSSTS 604
QY 1006 FOAHTPAOKASVLAFLINELACSKSVSEIDKNIDMNSLRDRKVV----- 1052
Db 605 VYDLTPGEKWIHLALCGKLL---TLVSTRDFIEDYVDILROAKQEFRELKAEQHRKERE 661
QY 1053 -----EKLRL-----KLRIHAKTKGRDTSGLDGERBQHPGLTPTPGKRKR 1096
Db 662 EAAAIRKRKEKLEQEQKMEKQEKLEKEDQORNSTADISIGEE-----R 708
QY 1097 KCGSDYDDDDDDDDDDDEDEDEED-----KEDQKKG-----KT 1133
Db 709 EDFDTSIESKDEQEKLQDMVTEDEDDPGSHKRGRRGKQNGQKFEFTROEQINCVTR 768
QY 1134 DICEDEDEDOQAASVEELEKQIEKLSKOOSQYRRKLFDAHSLRSMVFGPDYRRRYWIL 1193
Db 769 PLTADBE---ALKQHQKKELEKLEKIQS-----AIACNTNIFPLGRDMRYRWIF 817
QY 1194 PRGGIFVFGMESGEGLEIEAKERKLEKAEVQIKEMFETSGDLSNCTNDHCEQKED 1253
Db 818 PSIPGLFIE--EDYSGLTE----- 834
QY 1254 LKEKNTNLFLOKPSFSKLSKLLVAKMPPSEVMTPKPNAGANGCTLSQNSGKSLG 1313
Db 835 -----DMLLPRPSFQ---NVQSDQPVSTKTGEP----- 862
QY 1314 SVQSTATOSNEKADSNLNTGSSGPGKFSYPLND---OLLKTLTEKNQWFSLLPRT 1370
Db 863 -LMSST-SNIDQ-----PRDSVQLPKPVHPKPNR-W----- 892
QY 1371 PCDDTSLTHADMSTASLVTPOSPPSPSPPTAPLGSAAQNPVGLNPFALSPQVKGCV 1430
Db 893 -CFYSSCRLDQLEAL----- 908
QY 1431 SMWGLQFCGWTGVVTSNIPPTLSVPSLGSGLGSEGNGSFLTSNVASSKSESPVPONE 1490
Db 909 -----NSRGHRESALKETLLQEKSR----- 928
QY 1491 KATSQAAPAAVEAK---PVDFFSPKPIPEMQGWRRIIDPEDLKALLKVLHLR-----G 1542
Db 929 ---ICAQLARFEEKHFSDKPOQDSKPYTSRGRSSNAYDPQMCAL-EKQELRLRDFLLD 985
QY 1543 IREKALQ-----KQIQKHLDYITQACLNKDVAIIELENEENOVTRDIVENNVSERQA 1596
Db 986 IEDRIYQGLTGAIKVTDH---IWSALESGRYELLS-EENKENGIIKTV--NEDVEEME 1039
QY 1597 MEMDLSVLQOVEDLERRVASASLQVKGWMCPEPASEREDLVYFEHKSFTKCKHGDGFT 1656
Db 1040 IDEQTKVVKDRLGIIKTETSTSTVNSTAQSVS----- 1074
QY 1657 GEDESSAHLERKSNPLDIAVTRLADLERNIERIEEDIAP-----GLRVWR 1704
Db 1075 ----SVHYL-----AMALFQIEQGIERRFLK--APLADSDSGRYKTVLDRWR 1117
QY 1705 RALSARSAAQVALCQIQLOKSIAWEKSIKMYVCOICRKGDNELLLICDCKGCHTYC 1764
Db 1118 ESSLSSASLSQVFLHSLTDRSVIWSKSIINARCKICRKGDAENWVLCDCDGRGHHTYC 1177
QY 1765 HRPKITTTIPDGMWFCFACIAKASGOTLKIKKLH--VKGKKTNESKKGKVVLTGDTG--- 1819

Db 1178 VRPKLTVPEDGWFCPCRPKQRRRLSSRQRPSLESDVEDSMGGDEVDGDEBEGQ 1237
QY 1820 -----DEB----- 1822
Db 1238 SEEBEYEQEDDSDQEBEEVSLPKRGPQVRLPVKTRGKLSSSFSRGGQQQEPGRYPSR 1297
QY 1823 -----SASTSSSL----- 1830
Db 1298 SQQSTPKTTSKTSRSLKINSAPPTBTKSLRIASRSTRSHGSHGLQADVFVELLSPRK 1357
QY 1831 KRGNKOLQK-----RKMEVNTSN-----LSKQESFTSVKPKRDSK-- 1868
Db 1358 RGRKRSANKTPENSPNFPNFRVIATKSEQSRSVNIASKLSLOESKRRCKRKQSPSPS 1417
QY 1869 -----DLALCSMLTETMETHEDAWPFLLPVNLKLVPGVKVKKPMDPS 1912
Db 1418 PVTLGRSSRGQGVHLSAFEQLVVLVRHDDSWPFLKLVSKVQVPYDIKKPIALN 1477
QY 1913 TIREKLSSGQYPNLETFALDVRLVDCNCTFNEDSDIGRAGHNNRKYF 1961
Db 1478 IIREKWNKCYKLASEFIDDIELMFSNCFEYNPRNTSEAKAGTRLQAF 1526
RESULT 17
Q8CFP4 PRELIMINARY; PRT; 131 AA.
AC Q8CFP4; 01-MAR-2003 (TRENBLrel. 23, Created)
DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)
DE CDNA sequence BC053917.
GN Name=BC053917;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_taxID=10090;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=EYE;
EX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.B., Feingold E.A., Grouse L.H., Berge J.G.,
RA Klausner R.D., Collins P.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bobak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Sanchez A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whitting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN (2)
RP SEQUENCE FROM N.A.
RC TISSUE=EYE;
RA Strausberg R.;
RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC042646; AAH42646.1; -.
DR HSSP; Q03330; 1E61.
DR MGI; 3039620; BC053917.
DR InterPro; IPR001487; Bromodomain.
DR Pfam; PF00439; Bromodomain; 1.
DR PRINTS; PR00503; BROMODOMAIN.
DR SMART; SM00297; BROMO; 1.

DR PROSITE; PS50014; BROMODOMAIN 2; 1.
SQ SEQUENCE 131 AA; 15208 MW; 74E0A75AFB2AFAD1 CRC64;

Query Match 6.3%; Score 639; DB 2; Length 131;
Best Local Similarity 90.1%; Pred. No. 2.3e-17;
Matches 118; Conservative 9; Mismatches 4; Indels 0; Gaps 0;

QY 1842 MEENTSINLSKQESFTSVKPKRDSKDALCSMLTETMETHEDAWPELLPVNKLVPGY 1901
DB 1 MEETSINLSKQESFTSVKPKRDSKDALCSMLTETMETHEDAWPELLPVNKLVPGY 60

QY 1902 KVKIKPMDFSTIREKLSSGGYPNLETFALDVLRLVDFNCETFNEDDSDIGRAGHNMRYK 1961
DB 61 KVKIKPMDFSTIREKLSSGGYPNLETFALDVLRLVDFNCETFNEDDSDIGRAGHNMRYK 120

QY 1962 EKKWTDTPFKVS 1972
DB 121 EKKWTDTPFKVS 131

RESULT 18
Q8CGH2 PRELIMINARY; PRT; 769 AA.
AC Q8CGH2;
DT 01-MAR-2003 (TremBLrel. 23, Created)
DT 01-MAR-2004 (TremBLrel. 23, Last sequence update)
DE Baza2a protein (Fragment).
GN Name=Baza2a;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBT_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CZECH II; TISSUE=Mammary tumor;
RX MEDLINE=22389257; PubMed=1247932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heide F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Raba S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bobak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grinwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences";
RL proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CZECH II; TISSUE=Mammary tumor;
RA Strausberg R.;
RI Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC038012; AAH38012.1; -;
DR MGD; MGI:2151152; Baza2a.
DR GO; GO:0006338; P:chromatin remodeling; TAS.
DR InterPro; IPR000637; A:T hook.
DR Pfam; PF02178; A:T hook; 2.
DR Pfam; PF01429; MBD; 1.
DR SMART; SM00384; A:T hook; 2.
DR SMART; SM00391; MBD; 1.
FT NON TER 769
SQ SEQUENCE 769 AA; 83306 MW; 9507B9EEB30F4E1 CRC64;

Query Match 5.5%; Score 562; DB 2; Length 769;
Best Local Similarity 25.3%; Pred. No. 1.4e-13;
Matches 199; Conservative 119; Mismatches 246; Indels 224; Gaps 28;

QY 183 GLSVNKPPLSLVNOAKK---ETYMKLIVSPDVLKAGNKNTSESSLLTSELASKRBOYK 239
DB 30 GLYNGSPMFPQOQSLNGDVNNGLSVTSHTTTSGILNSAPHSS-STSHLHPNVAYD 88

QY 240 ---QAFPQLKQESSKSLKVIAALSNPKATSSSPAIHPKQTLNNHNPFLTNALLG 294
DB 89 CLWNSQYPS-----ANPGNNLKDPPLLSQPPGQYF---LNGILGG 127

QY 295 NHOP-----NGVIOSTQIE-----APLALT--TKTKMOSKINENIAAASSTP---FSS 337
DB 128 NROPSSPSHTNLRAGSQEFWANGTOSPMGLNFDSDQLYDSDPQNFVMPNGPPSFTS 187

QY 338 PVNLSTSGRTPGNQT--PVMPASPIILHSQGKEKAVSN----- 374
DB 188 P---QTSPLMGSSIQTFAPSDQVSSDIHPDEAAAEKELTSVVAENGTGLVGLSELEERPE 244

QY 375 -----NVNPKVTOHS-----HPAKSLVQFRGTDSIDIPSSKD-----SEDSNEDE 415
DB 245 LKMGYNGSVSVESLHQEVSVLVDPDPTVSCLDL-----PSHLPDQLEDTPILSEDSLEPF 300

QY 416 EE-----DDEB-----EDEDEDEDE-----SDDSQSE 438
DB 301 DSLAAABPVSGSLYGDADLGAEDKPLEGNPVISALDCPALSNNANASLLADDSQTS 360

QY 439 SD-----SNSESDTEGSEEDDDDDKQDQSDSDTEGKTSMKL 476
DB 361 ASIFVSPTSPPVLGESVLQDNSFGLNSCSDSEQEEIETQSSNFQRLTEPAPDQPPSIQL 420

QY 477 NKTTSVKSPMSLTGHS-----TPRNLIHAKAGSAPA--ALCSQSQ----- 517
DB 421 HPAVSPITASPAASLTASAEISPAVSPVSPVPEVFAVSPASSPALPAISLEASMTTP 480

QY 518 -----SPA-----FLGTSSSTLTSSPH-----SGTSKRRRYTDE 546
DB 481 VTSFQSGPEPSAAAFQTVSPASKNVSSAPKAAADRETTGGAVAGVSGSDVLKRRITATP 540

QY 547 RELRIPLEYGQWQRETRIRNFGRLQGEVAYYAPCGKKLQYQYEVIKYLSRNGIMDI SRDN 606
DB 541 EEVRLPLQHCWRRREVIRKSGSHWQGETWYFGCGKEMKQFPEVIKYSRNVHVSRRH 600

QY 607 FFSFAKIRVGDVFBEARDGPOEMQWCLLKEDVTPRTRAMEGRGRGRPPNPRQARESRM 666
DB 601 FFSFSPMPVGDFFFEEDTPGELQWVQLSABEIPSRQAITGKGRPRNNKAKNKEVPKV 660

QY 667 RRRKGRPPNVGNAEFLDNADAKLLRLQAOEIAQAAQIKLLRLQOEQARVAKEAKQ 726
DB 661 KRGRGRPPKIMPELLNKTNRPLPKLETQELISEDDKAKTKNKKWQO---KVQGEN 717

QY 727 QAINAAEEKQKQEQIKIMQOEKIKRIQOIRMEKELRAQOQILEAKKKKKEEANAALKE 786
DB 718 QTPVQGGQARNKRQDTKSLQKQTKKK--LKAKE-----KMKTKQE-----KLKE 761

QY 787 AEKRIKKEK 794
DB 762 KVKRKKKK 769

RESULT 19
Q96013 PRELIMINARY; PRT; 683 AA.
AC Q96013;
DT 01-DEC-2001 (TremBLrel. 19, Created)
DT 01-DEC-2001 (TremBLrel. 19, Last sequence update)
DT 25-OCT-2004 (TremBLrel. 28, Last annotation update)
DE LP06732p (CGI0897-FC).
GN Name=tau; ORFNames=CGI0897;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

FT CONFLICT 22 22 L -> P (in Ref. 3).
FT CONFLICT 136 136 K -> E (in Ref. 1).
FT CONFLICT 191 191 N -> D (in Ref. 4).
FT CONFLICT 298 298 Y -> V (in Ref. 3).
FT CONFLICT 823 823 E -> R (in Ref. 3).
FT CONFLICT 1191 1191 R -> P (in Ref. 3).
FT CONFLICT 1354 1354 K -> M (in Ref. 2).
FT CONFLICT 1438 1438 A -> V (in Ref. 3).
SQ SEQUENCE 1483 AA; 170902 MW; OCC146FEBB954261 CRC64;

Query Match 5.3%; Score 536; DB 1; Length 1483;
Best Local Similarity 18.0%; Pred. No. 2.8e-12;
Matches 348; Conservative 254; Mismatches 584; Indels 744; Gaps 66;

QY 221 SESSLLTSELRSKREY-----KQAPPQOLKQESSKSLKVVIAALSNPKATSSSPA 273
DB 28 TQEAFTREEEARLERYSERIWTCKSTGSLTHKEAWEERQEAELLK-----EEFPA 82
QY 274 -HPKQTLNHNPNFLTWALLGNHQPNGVIOEQAPLALTTKTMOSKINENTAAASS 332
DB 83 WYKCLVLEWHHN-----TASLEKLVDTAWLEINTKYAVGECDFEV-----124
QY 333 TPFSSPVNLSTSGRRTPGNQTPVMPSPASPILHSQCKEKAVSNVNPVKTQHHSHPAKSIV 392
DB 125 -----GKEKML-----KVIVKI-----HPL-----140
QY 393 EQFRGTDSDIPSSKSDSEDSNEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDE 452
DB 141 -----EKVDEATEKKSDGACDSSPSDKENSQIAQHQKKE 178
QY 453 EDDDDKQDQSDSD-----TEGE-----KTSMKL-----NKTSSVKS 485
DB 179 VYKDEGRESINDRARSPRKLPTSLKKGERKWAPPFLPHKYDVKLQNEDKIISNV-- 236
QY 486 PMSLTGHTSPNLHIAK-----APGSAPAALCSSESQSPA-----520
DB 237 PADSLIRTPPNKEIVFYFRHNALRAGTGENAPWVEDELVKYISLPSKFSDFLLDPY 296
QY 521 -PLGTSSSTL---TSSPHSGTSKRRVTDRELRIPLBYG-WQRETRIRNFG---RLQG 572
DB 297 KYMTLNPSTKRKNTGSPDRKPSKSK-TDNSSLSPLNPKLWCHVHLKKSLSGSLPKVKN 355
QY 573 EVAYYAPCGKLRQYPEVVKYLSRNGIMDI SRDNFSPSAKIRVGDFYBARQGPQMOWCL 632
DB 356 SKNSKSP---EEHLEEMKWMSPN-----KLHT-NFHIPKKGP-----389
QY 633 LKEEDVIRIRAMEGRGPPNPQRAREESRRRRKGRPNVGNCAEFLDNADAKLRK 692
DB 390 -----PAKPKGHSKPLKAKGRSGI-----411
QY 693 LQAQBIARQAQIKLRLKLOQEQARVAKEAKKQQAIIAABEKRKQEQIKIMKQOEKIK 752
DB 412 -----LNGOKSTGNSKSPK--GLKTPTKMKQMTLLDVAKGTQKMT 451
QY 753 RIQQ-----IRMEKEL--RAQOILEAKKKKKE-----AANAKLLEAEKR 790
DB 452 RAPRNSGGTPRTSSKPHKHLPPAALHLYAYKENKQREDKRSALSVCVISTARTLLSSEDR 511
QY 791 IK-EKEMR---RQOAVILKHOER-----ERRQHMMLKAMEARKAEKELKQEKDE 841
DB 512 ARLPPEELSLVQKRYELHEKRWASMBEQREYLLKKKREELKKLKEKAKERREKEML 571
QY 842 KRLNKERLEORRLEMAKELKKPNEDWCLADQKPLPELPRIGLVLSGTSFSDCLMVV 901
DB 572 ERLEKQKYEQEL---TGKML-----PAPFLVDTEB-GLPNTLFGDVANVV 614
QY 902 QFLRNFGLVGFVDVNIIDVNIISVLQEGLLNTGDSNGEVQDILLRLLSAVCDPGLITGYK 961
DB 615 EFLSCYGLLLPDAQYPIYAVSLMEALSADKGGFL-YLNRVLVILLQTLQDE-IAEDY- 671
QY 962 AKTALGEHLNVGNRDNVSEILQIFM-----FAHCQOTELT 998
DB 672 --GELGMLKSIPLTLHVSLSVLRLCLRRSDVQEESESGSDTDNDKDSAAFEDNEVQDEFL 729

QY 999 ESLTKTAFQANTPAQKASVLAFLINELACSKSVSEIDKNIDYMSNLRDRKVV-----EG 1054
DB 730 EKLETSEFFELTSBEKLIQTALCHRLIMTYSVQDHMETROQMSAELWKRILAVLKEEND 789
QY 1055 KLRKLRIIHAKTGKRTSGGIDILGEEQHPGLGTPPGKRRRKGGSDYDDDDDDSD- 1113
DB 790 KRAEKQKRKEMEAKNKENGKENG-----LG-----KTRKKKIVKFEPOVDTEABDM 838
QY 1114 -----QGDDEDEDEDEKQKTKTDICEDDEDEGQQAASVELEKQIEKLSQ 1161
DB 839 ISAVKSRLLAIQAKEREIQE--REMKVKLERQAEERIRKHKAARK-----885
QY 1162 QSOYRKLFDASHLSRSMFGPDYRRYRIL-PRCGGIFVEGMESSGLEEIAKEREL 1220
DB 886 --AFQEGIAKAKLVNRRTPIGTRDNRHNYLWFSDEVPLGFIE-----925
QY 1221 KKAESVOIKEEMPETSG---DSLNCSTNDHCEQKEDLKEKONTNLFLOKPSFKLSKLL 1277
DB 926 -----KGVVHDSIDYRFNHHC-----KQHT---VSGDEDYCPRSK-- 957
QY 1278 EVAKMPPESEVMTPKPNAGANGCTLSYQNSGKHSLSGVSQSTATOSNVKADSNNLFNTGS 1337
DB 958 -----KANLGKVA-----SWNTOHGTAATEVAVE-----980
QY 1338 SPGKFGYSLPNDQLLTKLTEKNRQ--WFSLLPRTPCDDTSLTHADMSTASLVTPQSOPP 1395
DB 981 -----TTTPKQGNLWFL-----CD-----995
QY 1396 SKSPSPAPILGSSAQNPNVGLNPFALSPLOKGVGVMGLQFCGWPTGVVTSNIPFTLSV 1455
DB 996 -----1456 PSLGSLGLSGNGNSFLTSNVASSKSESPPVQNEKATSAQAAVEAVKPVDFPSPKPIP 1515
DB 996 -----1516 EEMQFGWRIIDPEDLKALLKVLHLRGIREKALQKIQK-HLDYITOACLKNK-DVAIIE 1573
DB 996 -----SQKELDELLNCLHPQGIRESQLEKRYQDIHSHLARKPNLGLKS 1044
QY 1574 LNEHEE--NQVTRDIVENWSVEEQAMEMDLSVLOQVEDLERRVAS-----ASLQ 1620
DB 1045 CDGNQELNFRSLDIE---VATRLQKGLGYVEETSEFEARVISLEKLDKDFGECVIALQ 1101
QY 1621 -----VKGMWCPASEREDLVYFEHKSFTKLCKEHDGFTGEDESSAHALERKSDNP 1673
DB 1102 ASVKKKLGGMAPKQ-----KRRKLQS 1124
QY 1674 LDIAVTRLADLERNIERRIEEDIAPLRVMRRLSARSAAQVALCICQLOKSIKAWKSI 1733
DB 1125 EDSAKTEEVDEEKKWVE--EAKVASALEKWKTAIREAQTFSRMHVLLGLMDACIKWDMSA 1182
QY 1734 MKVYCOICRKGDNELLLLCDGCDKGCHTYCHRPKITTIIPGDWPCFACIAKASQOTLKI 1793
DB 1183 ENARKCVCRKKGEDDKLILCDKCNKAFHLFCLRPALYVDPGEWQCPACQAPATA-----1236
QY 1794 KKLHVKGKTNESKKGKVTILTGOTEDDS-----1823
DB 1237 -RRNSRGRNYTESASE-----DSEDDSEDESEEBEEDYEVAGLRLRPKRTIR 1289
QY 1824 -----ASTSSSLKRGNKOLQKREMEENTS--INLSKQESFTSVKPKRDDSODLALCS 1874
DB 1290 GKHSVIPPAARSRRPGKKPHSTRSPKAPKPVDDAEVDELVLQTKRSTRRSQSLQKCE 1349
QY 1875 MILTEMETHEDAWFLPLPVNLKLVPGYKVIKKPMDFSTIREKLSGGGYPNLETFALDVR 1934
DB 1350 EILHKIVKYSWFPREPVTREADEADYDVITHPMDFQTVQNKCGSGSYRSVQBFLEDMK 1409
QY 1935 LVFDCNCTFN 1944
DB 1410 QVFTNAEYVN 1419

```
RESULT 21
Q8UVR4
ID Q8UVR4 PRELIMINARY; PRT; 1079 AA.
AC Q8UVR4;
DT 01-MAR-2002 (TremBLrel. 20, Created)
DT 01-MAR-2002 (TremBLrel. 20, Last sequence update)
DT 01-MAR-2004 (TremBLrel. 26, Last annotation update)
DE Williams syndrome transcription factor (Fragment).
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=2166941; PubMed=11809820;
RA MacCallum D.E., Losada A., Kobayashi R., Hirano T.;
RT "ISWI remodeling complexes in Xenopus egg extracts: identification as
RT major chromosomal components that are regulated by INCENP-aurora B.";
RL Mol. Biol. Cell 13:25-39(2002).
DR EMBL; AF412333; AAL60161.1; -.
DR HSRF; Q9UIG0; IF62.
DR GO; GO:0000151; C:ubiquitin ligase complex; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0004842; F:ubiquitin-protein ligase activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0016567; P:protein ubiquitination; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR Pfam; PF00439; Bromodomain; 1.
DR Pfam; PF00628; PHD; 1.
DR PRINTS; PR00503; BROMODOMAIN.
DR SMART; SM00297; BROMO; 1.
DR SMART; SM00249; PHD; 1.
DR SMART; SM00184; RING; 1.
DR PROSITE; PS50014; BROMODOMAIN_2; 1.
DR PROSITE; PS50827; DDT; 1.
DR PROSITE; PS01359; ZF_PHD_1; 1.
DR PROSITE; PS50016; ZF_PHD_2; 1.
DR PROSITE; PS50089; ZF_RING_2; 1.
FT NON TER 1
SQ SEQUENCE 1079 AA; 124503 MW; 2AF4689FFE9E83DF.CRC64;

Query Match 5.2%; Score 527; DB 2: Length 1079;
Best Local Similarity 19.3%; Pred. No. 4.3e-12;
Matches 267; Conservative 218; Mismatches 449; Indels 446; Gaps 45;

Qy 670 KGRPNVGNABFLDNADAKLRQLQAOIARQAAQIKILRLKLOEQEQAARVAKEAKQQAII 729
Db 15 KKKPKN-GKSQVLNG--QKVGSKTRSPKKAIVKSPKPKQMTLLDMAKSTPKVSRQKGGSS 71

Qy 730 MAAEKRRKQK-----EQIKMKQOEKIKRIQIRMEKELRAQQOILEAKKKKKEEA 779
Db 72 TPRSCKPHKYLPAPALHLISYFRDNKNRDRKKSALISIKVAR--MLTAEDRKRLPD 128

Qy 780 ANAKLLEAEKRIKEKEWERQOAVLLKHOERERRRQHMMLKMAEARK-KAEKEKRLKQEK 838
Db 129 DLQELV--QKRYDLEHKKQWAVTEQOREEYMR-----KGRALKVRIKEKARERKQK 180

Qy 839 RDEKLNKERKLEQRRLLEMAKELKPKNEDMCWLADQKPLPELPRIPLGLVLSGTSFSDCL 898
Db 181 EREERLEKQRYEQEITGKSLPTFK-----LVD---TPF-----GLPNALFGDVA 223

Qy 999 MVQFLNFGKVLGFDVNDIVPNISVLQGLLNTGDSNGEVQDILLVRLLSAAVCDPGLIT 958
Db 224 MVIEFLSGYDILLPLDAQYPTAVSLMEALAAEKGGFM-YLNRVLVLVLLQTLQDE-IAE 281

Qy 959 GYKAKTALGELHLNVGNRDNVSEILQIFM---EAKCGQTELT----- 998
Db 282 DY---GELGMKLSIFPLVHLSASELVRICLRKSDSPAGESDSTKGDSDSVGSVAYQODE 338
Qy 999 -----ESLTKTKAQAMTPAQKASVLAFLINELACSKSVSVSEIDKNIDYMSNLRDRKVV 1052
```

```
Db 339 VEDEYLEKLETSEFFELTTBEKHLILAALCHRIILMTYSVODHVDA-----KQOR 387
Qy 1053 EGKLRKLRIIHAKTGKRDTSGGIDLGEQHPGLTPTPGKRRKRGSDSDYDDDDDDSD 1112
Db 388 SGELWKERLAIK----- 400
Qy 1113 DOGDEDEDEDEKEDQKKTIDICEDEDEGQAASVEELEKQIEKLSKQSQYRKLFPDA 1172
Db 401 -----EENDKKRAEKQKKEQGVKPEE-VQAQKI--VKKQ-EKINTQDNDADAEIMISA 451
Qy 1173 SHLSRVSVMFPDRYRRYRWWILPRCGGIFVGMESGEGLIEIAKEREKLKKAESVOIKEM 1232
Db 452 VKSR-----LQAMQA-----KKEKEHEKLTKEIRIERE- 480
Qy 1233 FETSGDSLNCNTDHCQEDLKEKONTNLFLOKPGSFSKLSKLEVAKMPPESEVWTPK 1292
Db 481 -----TEEERSRKQK----- 490
Qy 1293 PNAGANGCTLSYQNSGKHSLSGVQSTATQSNVEKADNNLNTGSSGPGKFYSLPNDQL 1352
Db 491 -----ASAeka----- 496
Qy 1353 LKLTLEKNRWFSLPRTCPDTSLTADMTSTASLVTPOSPSPSPPTAPIGSSAQN 1412
Db 497 -----PHEGIAKAKVLRRS-----PLGTDNRH 519
Qy 1413 PVGLNPALSPLOVKGVSVMGLQFCGPTGVVTSNIPFTLSVPSLSGLGLSEG---NG 1469
Db 520 ----NRYWL-----FSDVP-----GLYIEKGWHDSS 542
Qy 1470 NSFLTSNVAGSKSPVPONEKATSAQAAVEAVAKVPDFFSPKPIPEMOPGMMRIIDPE 1529
Db 543 INTRFSPESKQDSEDAESEDANSSIGCPDDSTQREKHAETTVPGQGNLWFLCDTQK 602
Qy 1530 DLKALLKVLHLRGIREKALQKIQ-KHLDYITQACLNKXQVAILIENENES---NQVTRD 1585
Db 603 ELDELLSLHPQGFRESQKLERLQNRQYQDQIMHSLHARKQNLGLKTCGQOQELNLFRLSD 662
Qy 1586 IVENWSVEEQAMENDLSVLOQVEDLERRVAS-----ASLQ-----VKGWMC 1627
Db 663 IIE---VATRLQKGLGLYLDTTTEFEAKVRTFENLKDGFECIVFLQAAVINKFLOGFMAP 719
Qy 1628 EPASEREDLVVFEHKSFTKLCKEHGDFGTGEDSSAHALKERKSDNPLDIAVTRLADLERN 1687
Db 720 KQKRR-----HQs-----EAAAKAAEQDEK-----KAAE----- 746
Qy 1688 IERRIEDIAPGLRVWRRLSEARSAQAVALCIQLOKLSIAWEKSIKMYVYQICIRKGDNE 1747
Db 747 ----EAKVASAVEKWKVAIRDQAQTFSRMHVLLGMLDACIKWDMSENARCKVCKKGED 801
Qy 1748 ELLLLCDGCDKGCHTYCHRPKITIIPGDWFCPAC---IAKAS---GQTLKIKLHVKGKK 1802
Db 802 DKLLTCDENKAPHLFCILRPVLPNIPDGEWLCFACQATARRSRGRNYAEDSTQDDEE 861
Qy 1803 TNESKKGKVTLTGDTDEDESASTSSSL-----KRGNKDLQKQKVENTSNLSKQ- 1853
Db 862 ERESESESESESESESESESESESESESESESESESESESESESESESESESESESE 921
Qy 1854 -----ESFTSVKXP-KRDDS KDIALCSMILTEMETHEDAWPFLPLVNLKLVPGYKK 1903
Db 922 YVEDTEADVEEMVRSQKPTSRQNOBQKCEEILAKLIKVRFSWPPFPFNADEIEDYTK 981
Qy 1904 VIKKPMDFSTIREKLSGGQPNLETFAVDVRLVLPDNCETNEDSDSDIGRAGHNWKYFEK 1963
Db 982 VVTPMDFOTMQSKCSGSGYQTVQEFNLNDLKLFGNTELYE-----AGSSQLSCLK 1034

RESULT 22
Q7PRP9
ID Q7PRP9 PRELIMINARY; PRT; 1357 AA.
AC Q7PRP9;
DT 01-MAR-2004 (TremBLrel. 26, Created)
DT 01-MAR-2004 (TremBLrel. 26, Last sequence update)
```


Db 715 -----KAKHSLNKENQEHRLN-----GSAEPMDVESDSPE----- 745
Qy 1357 TEKNQWFSLLRPTPCDDTSLTHADMSLASLVTPOQPPSKSPSPPTAPLGSQAQNPVGL 1416
Db 746 -----APTHFELLMCS----- 756
Qy 1417 NPPALSPLOVKGSGVMGLOFCGWPVGVTNSNIPFTLSVPSLGSGLSGNCSPLTNS 1476
Db 757 -----GDKRSCIVHD 766
Qy 1477 VASSKSESPVQNEKATSAQPAAVEVAKVPDPSPKPIPEEMQFGWRIIDPEDLKALK 1536
Db 767 SRNGQQR-----WAYIYKAEIDELIK 789
Qy 1537 VLHLRGIRE-----KALQKQIQKH-----LDVITQACLNKQDVATIELNENBE--- 1579
Db 790 ALNPNGRLREVELLIQELSLVRLSLIEQHAQKCPVDLLSLENETMRKFKMAWESETRKYGE 849
Qy 1580 -----NOVTR-DIVEN-WSVEEQAMENDLSVLOQVEDLER-----RVASA 1617
Db 850 ANFGLPNGTDLNEVMRLHLVDRIIQFENDIYTGDLGRL-KVKDMEKWRSDLLGNGYDAQC 908
Qy 1618 SLQVKQWMCPEPASEREDLVYPHKSFYTKLCKEHGFEFTG----- 1657
Db 909 KLO---W---GPGCKLEDEAGSDNESH-ETHEEDDGLALLOKVKARYPDGMYLAASADT 961
Qy 1658 -----EDSSAHALERKSDNPLDIAY-----TRLADLERNIERRIEB----- 1695
Db 962 KPLPDSDEEDQHT-----NAVSIPIAVHNMASALLQVEALGKRFLEKPEYGMKKWDPK 1015
Qy 1696 -----IAPGLVWRRLSEARSAQVALCTQOLQKSTAEWKSIMKVYCOQCRKGDNEE 1748
Db 1016 QEALKLACDSRLHWEVSLMESTSFAQVFLHLNLHDCIQWRRSTWKSLSCKVCRGRSDPE 1075
Qy 1749 LLLLCDCGCKGCHYCHRPKITTPDGMFCPCACIAK---ASGOTLKIKLHVKGKK--- 1802
Db 1076 KMLLDCENAGTHMFCFKPLRLSVPPGNWYCNDCVSLGSLNGQNEKDKKQATKKRKF 1135
Qy 1803 -----TNESKGGKK-VLTGT-----DTEDEDSASTSSSLKRGK----- 1834
Db 1136 VEEEDDEATDEEEKKDDMTDEDAEHENKHEDEVEDDESVTSPSSRVNGRLRPP 1195
Qy 1835 -----KDLQKRKNEENTSINLSKQESPTS----- 1858
Db 1196 RTRPTSRRLTSKEIEHQAQEDVSDGSDASLTAGEDTIEDSDEKVCQKCFYDGGEI 1255
Qy 1859 -----VKKPKRQD----- 1867
Db 1256 KCVQCKLFFHLECVHLKRPRTDFVCKTKPMPQRPRRHNSNGDHDDEBPKAKRPP 1315
Qy 1868 KDIALC-----SMILTEMETH 1883
Db 1316 NSLRICIDKTARPSNGNNNNNNNNSSVNNNNHRSRGRTEHMPNLNSAALYDLLEQIMKH 1375
Qy 1884 EDAPWPLLPVNLKLVGYKKVKKPMDFSTIREKLSGQYPNLETPALDRLVDFNCFTF 1943
Db 1376 KAAWPLRPLVLTSEVPDYHQIITKTPMDLAKIKSLNMGAYQLNEELSLDQLVFRNCIDLY 1435
Qy 1944 NEDSDSIGHAGHNMRY 1960
Db 1436 NVEGNEIYDAGCQLERF 1452

RESULT 26
Q9V9T4 ID Q9V9T4 PRELIMINARY; PRT; 1476 AA.
AC Q9V9T4;
DT 01-MAY-2000 (TEMBLrel. 13, Created)
DT 01-MAY-2000 (TEMBLrel. 13, Last sequence update)
DT 01-MAR-2004 (TEMBLrel. 26, Last annotation update)
DE CG1966-PA.
GN NamesAcfl; ORFNames=CG1966;
OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vandal M.D., Zhang Q., Chen L.X.,
RA Branton R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Abell J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Bessan P.V., Bereman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Gaig N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong P., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA Yeh R.F., Zavari J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426065; PubMed=12537568;
RA Celniker S.E., Wheeler D.A., Kronmiller B., Carlson J.W., Halpern A.,
RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,
RA George R.A., Hoskins R.A., Lavery T., Muzny D.M., Nelson C.R.,
RA Pacleb J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,
RA Svirskas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
RA Weinstock G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.;
RT "Finishing a whole-genome shotgun: Release 3 of the Drosophila
melanogaster euchromatic genome sequence."
RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426070; PubMed=12537573;
RA Kaminker J.S., Bergman C.M., Kronmiller B., Carlson J., Svirskas R.,
RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,
RA Ashburner M., Celniker S.E.;
RT "The transposable elements of the Drosophila melanogaster euchromatin:
a genomics perspective."
RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426069; PubMed=12537572;
RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Birman B.P.,
RA Bettencourt B.R., Celniker S.E., de Grey A.D., Drysdale R.A.,

RA	Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,	Qy	1082	-----QHPLGTPTRKERRKGGSDYDDDDDDDDQDQDDEDEDEKQKKTDTIC	1136
RA	Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,	Db	553	QTAQARQDLRALVVGKRLAA-----	586
RA	Lewis S.E.;	Qy	1137	EDEDEG-----DOAASVELEKQIKLSKQSQYRRKLPFADSHSLRS	1181
RT	"Annotation of the Drosophila melanogaster euchromatic genome: a	Db	587	HLEVNGVEPEKREALVEKLKKSIAELHAQSDQHRK-----HELQMLKLSHSLFNLVYL	641
RT	systematic review."	Qy	1182	GPDRYRRRYWILPRCGGIFVGMESGEGLEIAKEREKLAESVQIKKEPFETSGDSLN	1241
RL	Genome Biol. 3:RESEARCH0083-RESEARCH0083 (2002).	Db	642	GMDRCYRKYYVLESMPGIFV-----	668
RN	(5)	Qy	1242	CSNTDHCQEKEDLEKONTN---LFLQKPGSFKLSKLELV--AKMPPESEVMTPKPNAG	1296
RP	SEQUENCE FROM N.A.	Db	669	-----DTC-----LEQPIITNKSQIEIROQSALPKNRKDLRVLLKLYGDDEKTKK	714
RG	FlyBase;	Qy	1297	ANGCTLSYQNGSKHSLGVSQSTATQSNVEKADSNLNTGSSGPGKFYSPLPNDQLKTL	1356
RL	Submitted (SEP-2002) to the EMBL/GenBank/DBSJ databases.	Db	715	-----KAKHLENKENQEHLN-----GSAEPMDVESDPE-----	745
RP	SEQUENCE FROM N.A.	Qy	1357	TEKRWFSLLPRTPCDDTSITHADMTASLVTQSQPPSKSPSTPAPLGSSAONPVGL	1416
RG	FlyBase;	Db	746	-----APTHFELLMCS-----	756
RL	Submitted (MAR-2004) to the EMBL/GenBank/DBSJ databases.	Qy	1417	NPFALSPLOVKGVSMGLQFCGWPTGVVTSNIPTLSVPSILGSLGSEGNSEFLTSN	1476
DR	EMBL; AE003779; AAF57200.1; --	Db	757	-----GKRCIIVHD	766
DR	HSSP; Q9UIG0; IF62.	Qy	1477	VASSKSESPVQNEKATSAQAAVEAVKPVDFPSPKPIPEEMQFGWMRIIDPEDLKALK	1536
DR	FlyBase; FBgn0027620; Acf1.	Db	767	SRNGQRQR-----WAYIAEEIDELIK	789
DR	GO; GO:0016590; C:ACF complex; IDA.	Qy	1537	VHLRGIRE-----KALQKQIKQH-----LDYITQACLKQKDVAIILNENEE	1579
DR	GO; GO:0008623; C:chromatin accessibility complex; IPI.	Db	790	ALNPGLREYELLOELSVLRLIEQHAHTCPVDLLSLENETMRKFKMAAMESETNRKYGE	849
DR	GO; GO:0003677; F:DNA binding; TAS.	Qy	1580	-----NOVTR-DIVEN-WSVEQAMEMDLVLQOVEDLER-----RVASA	1617
DR	GO; GO:0042766; F:nucleosome mobilization; IDA.	Db	850	ANFGLPNTDLNENRHLVDRIIQFENDIYTGDLRL-KVKDMKWRSDLLGGNYDAQC	908
DR	GO; GO:0016584; F:nucleosome spacing; IDA.	Qy	1618	SLQVKGWMCBPASEREDLVYFEHKSFTKLCKEHDGEFTG-----	1657
DR	GO; GO:006355; P:regulation of transcription, DNA-dependent; IDA.	Db	909	KIQ-----W-----GPGKLEDEAGSDNESH-ETHEEDDGLLGYKARYKPYRDPGMYLAASADT	961
DR	InterPro; IPR001487; Bromodomain.	Qy	1658	-----EDSSAHLERKSDNPLDIIV-----TFLADLERNIERIEED-----	1695
DR	InterPro; IPR004022; DDT.	Db	962	KPLPDSDDDEEDQHT-----NAVSTPIAVHNMSALLOVEQAIGKRFLEPYGMKKWDPK	1015
DR	Pfam; PF00439; Bromodomain; 1.	Qy	1696	-----TAPGLRVWRRALSARSAAQVALCIQLOKSIKWEKSIKVVYCOICRKGDNBE	1748
DR	PRINTS; PR00503; BROMODOMAIN.	Db	1016	QKALKACDRLHQLHVEVSLMESTSPAQVFLHNLHDCIQWRRSTNKLKVCRRGSDPE	1075
DR	SMART; SM00571; DDT; 1.	Qy	1749	LILLCDGCKGCHTYCHRPKITTIIPDGWDFCPACIAK---ASGQTLKIKLHVKKK---	1802
DR	SMART; SM00249; PHD; 2.	Db	1076	KMLLDCENAGTHMFLKPLRSVPPGWNVCNDCVKSLGSLNGQNEKKQATKKRKF	1135
DR	PROSITE; PS50014; BROMODOMAIN_2; 1.	Qy	1803	-----TNSKKGKK-VTLTG-----DTEDEDSASTSSSLKRGV-----	1834
DR	PROSITE; PS00827; DDT; 1.	Db	1136	VEEDEDEATDEEBEKKDDMTDEDAEHENKHDDEDDDESVTSTPSSSRVNGRILRRP	1195
DR	PROSITE; PS01359; ZF_PHD 1; 2.	Qy	1835	-----KOLQKRWEEENTSINLSKQSFSTSVKKPKRDDS KDALCSM-----	1875
DR	PROSITE; PS50016; ZF_PHD 2; 1.	Db	1196	RTRPSTRRLTSKEIEEHAQVDSDGSDASLTAGEDTIEDSEDEKVCQKCFDGGEI	1255
SQ	SEQUENCE 1476 AA; 170410 MW; 9EE1906672E5B006 CRC64;	Qy	1876	-----	1875
	Query Match 4.7%; Score 481; DB 2; Length 1476;	Db	1256	KVCQRLFFHLECVHLKRPRTDFVCKTKCPQPRRRHNSNMNGDHRDEBEPKAKRPR	1315
	Best Local Similarity 18.9%; Pred. No. 3.3e-10;	Qy	1876	-----	1883
	Matches 310; Conservative 192; Mismatches 447; Indels 688; Gaps 55;	Db	1316	NSRLSIDKTARPSNGNNNNNNNNSSVNNNNHRRSGRETNEHMPNSAALYDLLSQIMKH	1375
Qy	665 RMRRKRGPPNVGNVNAEFLDNADAKLLRLKLAQAIARQAQIKLLRKLOKQEQARY----				
Db	163 RLRPNGKD-----SAB-LDLPPRQUR-----SRMEFNLENLNFYKSNVSRVDGLLR				
Qy	720 -AKEAKKQQAIAABE-----RKQKEQIKIMKQKEKIRIQIRMEKELRAQQAILEA				
Db	211 KPPEAYKVYVTPDGVNFSTIFIGKMPYSPAKIKKPDGK-----KQSTLNKYIVAGEATAA				
Qy	772 KKKKEEBAANAKLLAEKRIKEKVRQQAVALLKHOERRRQHMMLKAMEARKKAEK				
Db	267 KSAKAKSDAKSLAELEERVK-----REKEAKLIE-----LEKQAKKKAQLI				
Qy	832 ERLKQEKDEKRLNKKERLEQRLEMAKELKPNEDMCLADQKPLPELPRI PGLVLGG				
Db	310 ERVENE-----CNLLQKTDDLERTDQKVLPRYQIVTL-LPE				
Qy	892 STFSCLMVQPLRNFGKVLGPDVNDVFNLSVLQ- EGLNITGDSMGEVQDILLVRLLSAA				
Db	347 HLLGDAFWMRPFMHTYTGLS-GIEVFRQNLSFYEMTRALTARETAGPLSDILLVLL-GT				
Qy	951 VCDPGLITGYKATLGEHLLNVGNRDNVSIILQIFMEAHCGQTELTESLKT-----				
Db	405 VFD-----LQKEEECAVTLDRAAQTOEPYWSM				
Qy	1004 -KAFQAHTPAKASVLAFLINELACS KSVSVSIDK-----NIDYMNLRDRKVV-----				
Db	435 AQAAKSHLYAKGH--FSFKVNLPLDALTLSEVLRLHLGSGAFVNEKAERVRVYRNGY				
Qy	1053 -----EGKLRLRIIHAKTGK-----RDT-----SGGIDLGE-----				
Db	493 SKKDPGLRLHSHILRLKNSHVYQLKFKDILLIRCLMSQIMTYSGTINLIERME				

QY 1810 KK-VLTG-----DTEDEDSASTSSSLKRGN-----KDLO 1838
Db 1151 KDDDMTDEDAHENKEHDEVEDDESVTSSSRVNGRIILRPRTRPTSRRLTSKEIE |::: 1210
QY 1839 KRKMEENTSINLSKQESFTSVKKPKRDDSKDALCSM----- 1875
Db 1211 EHAQEDVSGDVSDASLTAGEDTIEDSDDEKVCQCFYDGGELKVCQRLFFHLECVH 1270
QY 1876 ----- 1875
Db 1271 LKRPRPTDVCCKTCKMPQPRRRRRHNMNGDHRDEEBPKAKPRNSRLSIDKTARPSN 1330
QY 1876 -----LLEWETHEDAWPELLPVLNKLIV 1898
Db 1331 GNNNNNNNSVNNNNHRRSRGRTEHNEHPLNSAALYDLEQIMKHKAAMPFLRPVLTVSEV 1390
QY 1899 PGYKVIKKPDMFSTIREKLSGGQPNLETFALDVRLVDFDCETFNEDSDSDBGAGHNR 1958
Db 1391 PDYHQIITPMDLAKIKSLNKGAYQLNEELLSDIQLVFRNCDLYNVESGNIYDAGCQLE 1450
QY 1959 KY 1960
Db 1451 RF 1452
RESULT 28
Q6P1D9
ID Q6P1D9 PRELIMINARY; PRT; 892 AA.
AC Q6P1D9;
DT 05-JUL-2004 (TRENBLrel. 27, Created)
DT 05-JUL-2004 (TRENBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TRENBLrel. 27, Last annotation update)
DE BC065123 protein (Fragment).
GN Name=BC065123;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Brain;
RX MEDLINE=23289257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Greenwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smalius D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Brain;
RA Strausberg R.;
RL Submitted (JAN-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC065123; AAH65123.1; -
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR001487; Bromodomain.
DR InterPro; IPR011011; FYVE_PHD_Znf.

DR InterPro; IPR001965; Znf PHD.
DR Pfam; PF00439; Bromodomain; 1.
DR Pfam; PF00628; PHD; 1. DOMAIN.
DR PRINTS; PR00503; BROMODOMAIN.
DR SMART; SM00297; PHD; 1.
DR SMART; SM00249; PHD; 1.
DR PROSITE; PS00633; BROMODOMAIN_1; 1.
DR PROSITE; PS00114; BROMODOMAIN_2; 1.
DR PROSITE; PS01359; ZF_PHD_1; 1.
DR PROSITE; PS00116; ZF_PHD_2; 1.
FT NON TER 1
SQ SEQUENCE 892 AA; 101830 MW; 95A6C798F37AA93F CRC64;
Query Match 4.7%; Score 474.5; DB 2; Length 892;
Best Local Similarity 19.6%; Pred. No. 3.3e-10; Indels 435; Gaps 35;
Matches 217; Conservative 144; Mismatches 310;
QY 1043 SNLRDRKVVVEGKLRKLRHIAKKTGKRDTSGGIDLGEEQHPLGTPTGKRRRRKGGSD 1102
Db 5 ARIRRK---EELKEQEQKMKKEQK-----LKEDEQRNSAAVFEGER-----ED 48
QY 1103 YDDDDDDSDQDGED-----DEDEED-----KEDQKQK--KTDI--CEDEDEGDAQASVEE 1150
Db 49 FDTSTENKNIEQKDLDPDVTTEDEDDPGSHKRSRRGKVGQTAVKQCIKQEEEMNYCIKQEP 108
QY 1151 LEKQIEK--LSKQOSQYRRKLPDASHSLRSVM---FGPDRYRRRYWILLPCGGIFVEGME 1205
Db 109 LSADAEALRQEQQKKEKLDKIOSAICTNIPFLGRDRLYRRYWIIPSPGLFIEDY 168
QY 1206 SGEGLIEIAKERKLLKKAESVQIKEMPETSGDSLNCISNTDHCQEKDLKEDKNTNLFQ 1265
Db 169 SG-----LTEMLLPPSPSFH--NNAQPRDPQVSIKTEES---FLS 204
QY 1266 KPGSFSKLSKLLVAKMPPESEVMTPKNAGACTLSYQNSKGXHSLSGVSTATQSWE 1325
Db 205 ESTS-----SLDQGFDDSVLLPKP-----VH 226
QY 1326 KADSNLNTGSSGPGK--FYSPLPN--DOLLKLTLEKNRQWFSLLPRTPCDDTSLTHADM 1382
Db 227 K-----PNRWCYSSCAQLDLDALNSRGHR--ESALKEVLLQEKSRICAOQL 272
QY 1383 STASLVTPSQSPKSPSPTPAPLGSSAQNPVGLNPFALSPLQKGVSMGLQFCGWPT 1442
Db 273 AHFS---EERHFDKQADSKPVSS----- 295
QY 1443 GVVTSTNPFTLSVPSLGLSGNGNSFUTSNVASSKSPVPQNEKATSAQPAAVEV 1502
Db 296 -----RCRSSGACDISQMSABRQLELR-- 319
QY 1503 AKPVDPPSPKPIPEMQFGWRIIDPED-----LKALLKVLHLRGIREKALQKIQKHL 1556
Db 320 ----DF-----LUDIEDRIYVQGTIGAI-- 346
QY 1557 DYITQACLKNDKVAIIELNENEENQVTRDIVENNSVEEQAMEMDLISLVQVEDLERRVAS 1616
Db 347 -----SALENGRYELLS--EESKENGVIKTV--NEDVEEMEMEQAARVIV----- 386
QY 1617 ASLVQKGMCEPASEREDLVYFHKGFTELCKEHDGFTGEDSSAHLERKSDNPLDI 1676
Db 387 -----RDRLLGIKTE-----TPSTISTASTPQSVSNVVHY 417
QY 1677 AVTRLADLERNIERRIEDIAFG-----LRVWRRLSEARSAAQVLCIQLOKS 1726
Db 418 LALALFOIEQGIERRFLKAPLDGNDSGRSYKTVLDWRRESLLSSASUSQVFLHSLTDRS 477
QY 1727 IAWKSIKMYVCOICRGDNEELLLCDGCKGCHTYCHRPKITTIPOGDFPCACIACA 1786
Db 478 VVMSKSIINARCKICRKKGDRENWVLCDGCDRGHTYCVRPKLKAVPDGDFCPCRPKQ 537
QY 1787 SGQTLKIKKLHVKKTKNESKKG-----KKVTLLFGTDEDEDSASTS 1827
Db 538 RSRLSSRQ--RPSLESDEEMEGMEDDDDDDEVDDEDEGQSEEEVEVEQDEEDSDDEAL 596


```

QY 1828 SSLKRG-----NKOLOKRX 1841
DB 597 SPKRGQVQLPIKTKGRFPSPRSORQDPRYPSSQOSTPKNTAKSASKNLRKTR 656
QY 1842-----1841
DB 657 SAPPTETRSIRVGSSTRHSPSALQDVVELLSPHSKRRGRKGADHTPEHSPSFTNFRVS 716
QY 1842-----MEENTSINLSKQESFTSVKPKRDRSK-----DLALCSM 1875
DB 717 TGRSSRLPLNTAESLSLQHSSEKRRGRKQSTESSPVPLNRRSSGRQGVHLSAFQ 776
QY 1876 ILTETHEDADPFLPLVNLKLVGKVKIKRPMDFSTTREKLSSQYPNLETALDVL 1935
DB 777 LVVELVRHDSVPFLKLVSKIQVPIYDIYIKKPIALNIIREKVNKCEYKLASEFIDIEL 836
QY 1936 VPDNCTFNEDDSDIGRAGHNRKYF 1961
DB 837 MFSNCFEYPRNTSEAKAGTRIQAFP 862

RESULT 29
Q75JP5
ID Q75JP5 PRELIMINARY; PRT; 1508 AA.
AC Q75JP5;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Similar to Babesia bigemina, 200 kDa antigen p200.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycota; Dictyosteliales; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AX4.
RX MEDLINE=22092622; PubMed=12097910; DOI=10.1038/nature00847;
RA Gloeckner G., Eichinger L., Szafiranski K., Fachebat J., Dear P.,
RA Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K.,
RA Tunggal B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.;
RT "Sequence and analysis of chromosome 2 of Dictyostelium discoideum.";
RL Nature 418:79-85 (2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=AX4;
RA Baumgart C.;
RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC116963; AAS45390.1; -.
DR HSSP; Q01082; 1AA2.
DR InterPro; IPR001715; Calponin-like.
DR Pfam; PF00307; CH; 1.
DR SMART; SM00033; CH; 1.
DR PROSITE; PS00021; CH; 1.
SQ SEQUENCE 1508 AA; 173964 MW; 3C8661F21658F2AB CRC64;

Query Match
Best Local Similarity 49.5%; Score 458.5; DB 2; Length 1508;
Matches 318; Conservative 279; Mismatches 595; Indels 415; Gaps 60;

QY 4 TKSTSSGGRNRKCNQEQSKNQPLDARVDKI KDKKPKKAM--ESSS-----NSDSD 52
DB 98 TTTTSSSSNQQLPYDEPPLVHVELTSKGRPPSPFKFESSQLDWTWINSDD 157
QY 53 SGTSSDTSSEGISSDSDLEED-----EEEDQSTSESDSDSDSESAHQHNNQV 105
DB 158 ENSKNDFKGFMSQLESSLNLEKYITSNIYQYVFPQKDSDDLFPSPPIKKQSNLD 217
QY 106 LHGISDPKADGQKAEKRIHQPLAFESQTHSPSQSQKQPVLSQQLPFIFOSS 165
DB 218 EKNIFDDDDDDIFKNKQSTVTKPKQP-----QQQKQPIV-----KK 255
QY 166 QAKESVNVKHTSVIOSTGLVSNV--PLSLVQNAKETYMKLIVSPDVLKAGNKNTSE 223
DB 256 KVEESLFGDSSLSDSLASKSPPKSVNL-----SPQLFK---EDSFDI 300

```

```

QY 224 SSLLTSELRSKREYKQAFPSQLKKQESSKSKKKVIAALSNPKATSSSPAHPKQTLENNH 283
DB 301 DSFLSND-----FDFKSKSNNTTTVKAV-----PISKSKQPPQKEIKE 343
QY 284 PNPFITNALLGNHQNQVIOQIAPLALTTKTKMQSKINENIAAASSTPSSPVNLST 343
DB 344 VSTQKDNSSNN--NGWVS--VDEPKPERST-SKSKIGIKQTVTVTKSNNSPENDIFGST 398
QY 344 SCRRTPG-NQTPVMPSPASPLHSQKKEKAVSNVNPVKTHHSHHPAKSLVQPRGTDSDI 402
DB 399 TTNDGGDNDFSTPATTPSSSSSTKATTS-----429
QY 403 PSSKDSSESDNE-EEDEDEDEED-----EDDESDDSQSESDSNSSDTEGSSEED 455
DB 430 PSTTTTKSNINIGKSNKSDQAQDLNDIFQEEQDKRREERAKLQKQKQKEQI 489
QY 456 DDKQDSESDTEGEKTSMLNK-TTSSVKFSMSLTGHTSPNHLHIAKAPASAPALCS 514
DB 490 KDEIDD-----LFKFSKPTTTTSTPSKSTTSIVDINSIEKKFSKSAANATTN 538
QY 515 ESQSPAFLTGTSSTLTSSPHSGTSKRRRVTDRELRIPLEYQWQRETRR-----NFGGR 569
DB 539 NDNN-----NNNNTSSP-----TNKSTNLFND-----WEKEEIKVKEELDKKK 579
QY 570 LQGEVAYYAPCGKLRQYREVIKYLSRNGIMDISRDNFSFSAKIRVDFYARDGQPMQ 629
DB 580 IQEQ-----KLEQEKULIEKKR--IAEKRISDEILAKQLAEKLEKRIEKELE 629
QY 630 WCLLKEEDVPIRAMEGRGRPPNPDPQRAEESRMR-----RKGPPNVGNNAFLD 683
DB 630 DLRLAKELEKRLALROEKELAEKLEKREKAEADKRIAQIEIKRL-----EKEQD 684
QY 684 NADAKLLRLQAEIARQAQIKLRLKQKQKQARVAKEAKKO-QALMAAEKRRKQEQI 742
DB 685 QLEKO--RLQEQRLQKQKDEKELADRLKEKEIENEIKQKQLEKIKLEKELAEKKE 742
QY 743 KIMQO-----QEKIKRIQIRMEKELRAQIILEAKKKKKEEANAANAKLLEAEKRIK-- 792
DB 743 RLQKEADEKRIADQLEFERLLKQKELAEK--LEKERLEKEAEEKRIAAEKLEKQR 800
QY 793 -EKENRQOAVLLKHQERRRRQHMMKAMEARKKAEKELKQEKR--DEKRLNKERKL 850
DB 801 LEKEAEKRIA---QDLERK---LEKEAEKRIADLERKLEKEAEKRIAAEKLU 852
QY 851 EQRRLLEEMAKE-LKKPNEDMCLADQKPLPELPIPLGLVLSGTSFSDCLMVVQFLRNFCK 909
DB 853 QQQELAAKLEKERLEKEAEKRIAQEKRIAEENRI-----887
QY 910 VLGFVDNIDVPNLSVLQEGLLNIGDSMGVEVDLLVRLLSAAVCDPGLITGYKAKTALGRH 969
DB 888 -----AQEKIAEB--LEKKELQEEQD-----RLAAAELEKRLKEAEKRIAQE 932
QY 970 LLNVGNRDNVSEILQIFMEAHCGGOTELTESLTKVAFQAHPTAQKASVLAFILNELACSK 1029
DB 933 LEKLEKE-AAEVKRIADEAAAAKLEKERLEKEAEKRI-ADEAAAAAKLEKERLEKE 990
QY 1030 SVVSEIDKNIDYMSNLRDKWVVEGKLRKRIIHAKTGKRTDTSGLDIEGEOHPLGTPT 1089
DB 991 AAAAE-EKRI-----ADEAAAAKLEKERLEKEK-----IADEAAAAEAAA 1031
QY 1090 PGRKRRRGSDYDDDDDDSDQDDEDEKEDKQKGGKTDICEDEGDQAAVSVE 1149
DB 1032 LLOQKIEK-----EKEERDIAKENKELKEKEDKERKEQORKEQERAEKRAALK-E 1083
QY 1150 ELEKQIEKLSQSQYRKLFDAHSLSRSMFDPDRYRRYRWWILPRCGGIFVEGMESGG 1209
DB 1084 KIEKERLNQKLDQEKEREERERERK-----EQEREEN 1120
QY 1210 LEETAKEREKLAESVQIKSEMPFTSGDSLNCSTNDHCEKEDLKEKXNTNLFLOKFGS 1269
DB 1121 EKQLEKEREKERLEKQORNEQLEKQE-----RFKKQOEKEK-----1161

```

QY 1270 FSKLSKLLVAKMPSEVMTPKPNAGANGCTLSYQNSGKHSGLGSVQSTATQSNVEKADS 1329
DB 1162 -----QLKEQQQKVILP-----TTTTTIRESNSDS 1190
QY 1330 NNLFWNTSGGPGKGFYSPL-----PNDQLLKLTKRQWPF-----SLLPRT 1370
DB 1191 DALLNLPDSASSSHSELVHLTLGRSKAKGRKKPTRBELTKDGNRNKQINPNLKLKPKST 1250
QY 1371 ---PCDDTSLTHADMS-----TASLVTPSQSPSKSP 1399
DB 1251 VIIPDDDDQSQQSQSEVEEKPVAVKKMPTGFNGMMMPKFPDPSAVKLKSAKPPSTNT 1310
QY 1400 SPTAPLGSAAQNPVGLNPFALSPLOVKGVSMMGLQFCGWPTGVVTVGNIPFTLSVPSLG 1459
DB 1311 TTTTIT---TAPQOPQVFN---LKPVALKPTVTO-----TTTTTTPPTT---PPS-- 1351
QY 1460 SGLGLSEG-----NGNSFLTNSVASKSESPPVQNEKATSAQPAAVE 1501
DB 1352 SSVQLKPAATRSFGSGSPFWGINSTTPTTPTTTPPSLKKSAKADGMK 1398
RESULT 30
Q9NSL9
ID Q9NSL9 PRELIMINARY; PRT; 1427 AA.
AC Q9NSL9;
DT 01-OCT-2000 (TREMELrel. 15, Created)
DT 01-JUN-2001 (TREMELrel. 17, Last sequence update)
DT 01-MAR-2004 (TREMELrel. 26, Last annotation update)
DE Hypothetical protein H20J04.2;
GN Name=H20J04.2; ORFNames=H20J04.2;
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoides;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_taxid=6239;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=99069613; PubMed=9851916;
RG WormBase Consortium;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018 (1998).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Dauphin S., Antoniou B., Gibson A.;
RT "The sequence of C. elegans fosmid H20J04.";
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
[3]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Waterston R.H.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
[4]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Waterston R.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
[5]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Waterston R.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
[6]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Waterston R.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
[7]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Waterston R.;
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.

RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Waterston R.;
RL Submitted (OCT-2002) to the EMBL/GenBank/DBJ databases.
[9]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Waterston R.;
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
[10]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RG WormBase Consortium;
RL Submitted (SEP-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC006661; AAF39888.2; -.
DR HSSP; Q03330; 1E61.
DR WormBase; WBGene00019217; H20J04.2.
DR WormPep; H20J04.2; CE27187.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR000637; A+T hook.
DR InterPro; IPR001487; Bromodomain.
DR InterPro; IPR004022; DDT.
DR InterPro; IPR001965; ZnF PHD.
DR Pfam; PF02178; AT_hook; I.
DR Pfam; PF00439; Bromodomain; 1.
DR Pfam; PF00628; PHD; 1.
DR PRINTS; PR00503; BROMODOMAIN.
DR SMART; SM00384; AT_hook; 1.
DR SMART; SM00297; BROMO; 1.
DR SMART; SM00571; DDT; 1.
DR SMART; SM00249; PHD; 1.
DR PROSITE; PS00633; BROMODOMAIN_1; 1.
DR PROSITE; PS50014; BROMODOMAIN_2; 1.
DR PROSITE; PS0827; DDT; 1.
DR PROSITE; PS01359; ZF_PHD_1; UNKNOWN_1.
DR PROSITE; PS50016; ZF_PHD_2; 1.
KW Hypothetical protein.
SQ SEQUENCE 1427 AA; 165227 MW; 401039F88C2BA4D3 CRC64;
Query Match 4.4%; Score 449.5; DB 2; Length 1427;
Best Local Similarity 18.5%; Pred. No. 5e-09;
Matches 305; Conservative 223; Mismatches 531; Indels 587; Gaps 64;
QY 515 ESQSPAFILGTSSTLTSSPHSGTSKR-----RVTDRELRIPLEYGWQRETRINFGG 568
DB 158 EEQNPSEFGSGSELKEPILPPDTFRYTLQILDPAVSEEMYREGIAF--DRLFRSKNIGS 215
QY 569 RLQGEVAYVAPCGKKLQYPEVIKYSRNGIMDISRDNFSAKIRVGDVEARDGPQEM 628
DB 216 R-----QKIR-----LFLKNCQSPDSERYTLKFQ---FLEKID---NL 249
QY 629 QW---CLLKEEDVIPRIAMEGRRGRPNP-----DRQARAEESRMRR-----KGRPP 674
DB 250 YWTDVMSGAEPICQTPALQ--RGAPNYTLXAGDHERKEKPKKEERDPAAPPRGRPP 307
QY 675 NVGNAEFLDNADAKLRLKLOQETARQAAQIKLRLKLOQEQARVAKAKQQAIAAAEE 734
DB 308 KTPQEKALDKKKEQMKRAEAG-----REAPKPDPDFLTSSGAMTSLSPSPA 355
QY 735 KRK-QKEQIKIMKQOEKIKRIQIIRMKEKELRAQQLLEAKKKKEAANAALLAEKRIKE 793
DB 356 KKKIRREGSGSMGKKTKILEQ-----QGELCFPF 386
QY 794 KEMRRQAVLLKHQERERRRRHMMMLKAMEARKKAEKERLKEKRDKRLNKRKLQJR 853
DB 387 SEARR-LGIDVGGLEQEE-----KLLSGKAVADFQRVKEEKDAE----KERAKEEK 433
QY 854 RLEEMAKELKKPNEDMCIAOKYLPPLPR--IPGLVLSGTSFSDCLMVVQFLRFGVL 911
DB 434 KRKREKTAYNKRGEDLINCCLKPLPRPKLEIPEWI-SNAEFEDYLFIFQPFNSFKOLL 492

QY 912 GF-----DVNLSV--LOEGLNIGDSMGVQD-----LL 943
Db 493 PLKEIRGSDEVPQSDIIIAIKCNDPQNSFADLLRVLLSIRTDAIEEDGDEADINREE 552
QY 944 VRLLSAADCPLGIT-----GYKAKTALGHELLNVGNVRNVSEILOIFNEA-- 990
Db 553 VYLINAQNCOPAHVTHGDSIRLSDLHFKIRKTHGKSVRHLPDVDMWTLTEVLRLIFETSG 612
QY 991 -----HCGQTELTESLKT-KAFOAHTPAOKASVL 1018
Db 613 YITGMATHRLHLYARNGFRGYEDPAYEFTRHPG---IMEKRLTLTVFLEAP-ERLEIV 668
QY 1019 AFLINELACSKSVVSEIDKNIDYMSNLRDK-----WVW--EGKRLKRLRIIHAKTGRD 1071
Db 669 KTLIYQLLYTSKPRGHLEQONELVELKREQKKAWDVQGEAANAARLL--LELAPAG 726
QY 1072 TSGGIDLGEOHPLGTPTPRKRRKRGSDYDDDDDDDDGDEDEDEDEDEKQK 1131
Db 727 ASGAPVKSQKEP---PV---VRLKA-----HIKAIINEGR 756
QY 1132 KTDICEDD-----EGDOAAVSELEKOIEKLSKQSOYR-----RKLFDASHSLRSYM 1180
Db 757 RYD-KEDDAIVLESVPYASL-SLDEIVTARELQKSEFKILMDSLISKMFQIYSKSIDIR 814
QY 1181 FQDRYRRYRWPILPRCGGIFVEG---MESGEGLEETIAKERKIKAAESVQIKEEMFETSG 1237
Db 815 LGSDRAYRYVIMENLSAILVETATFQELGIYCDPSIIDDDPLNLENEH--QEVFICGT 872
QY 1238 DSLNCNTHCQEKDELKEDNTNLFQKPGFSKLSKLELVAKMPESEVMTPRNAGA 1297
Db 873 -----NMDCRVHGDSENNWTRWSYIRDREOPEQLKSL-----NPRGNREV 914
QY 1298 NGCTLSYQNGKHSGLSGVOSTATQSNVEKADSN---NLFTGSSGPKFYSLPNDQLK 1354
Db 915 E--LLEELNEYRPSLEIIEETRLHEBEDEDEKWSQFMTNDPNPDY-----962
QY 1355 TUTEKNRWFSLPRTPCDDTSLTHADMSTASLVTPOSPQSPSPPTAPLGSSAQNVP 1414
Db 963 -----NIDW-----DAEMRDL-----974
QY 1415 GLNPALSPLOVKGVMGLQFCGWPTGVVTSNIPFTLSVPSGLSLSGNGNSFLT 1474
Db 975 -----LDF-----977
QY 1475 SNVASSKSPVQNEKATSAOAAVEVAKPVPFSPKPIPEMQFGWRIIDPEDLKAL 1534
Db 978 -----EKIDOGQMGST-----KIFECNRTEW-----RNLKES 1007
QY 1535 LKVLHLGIRKALQKIQKHLDIYTOACLKXKNDVAII--ELNENEENQVTRDIVENWSVE 1593
Db 1008 GNV-----CMLNEDISIFGEISVNLEES-----EHS-- 1035
QY 1594 EOAMEMDLVLQOVEDLERRVASLQVKGWMCPEPASERDLVTFEHSFKLCKEHDG 1653
Db 1036 -DSMKLAIAFYMIKSIHLKFIKAP-----YISPN-----KDEHG 1069
QY 1654 EFTGEDESSAHLERKSDNPLDIATVRLADLERNIERREEIDAPGLRWRRLSEARSA 1713
Db 1070 NL-----KPELFLR-----WQALLECESH 1090
QY 1714 AQVALCIOLOKSIKAWKSIKMYVQICRKGDNELLLCDGCKGCHTYCHRPKITITP 1773
Db 1091 SALSLSFISFTFEGSIKWDKSRLOQKCSRRKAAADHLVLCSECDNYHLKCAKLDVNSDA 1150
QY 1774 DGDWFCPACIAKAGOTLKIKLHVKGKTNKSGKVKTLTG-----DTEG 1820
Db 1151 PADWMTCTSCRAQQRKVENEAKMARDDDDQLAESSSQEDVAIGNISDNFSEIAPKTAEP 1210
QY 1821 EDASSTSSSLK-----RGNK-----DLQKRWK-----1843
Db 1211 AESCSTPTTSKEPPTIRTASGRAVKRVHVHSDIHEGLSLKQRKNSGSMPTPSPRVNVS 1270

QY 1844 -----ENTSI-----NLSKOESFTSVKK-----PKRDD-----SKOLALCS 1874
Db 1271 VRIFDVENELTDEDDSDGENTRKGKTKPTSKSVTSTPTTNDISRVIIPNIKKEKVTLIE 1330
QY 1875 MILTEWETHEDAWPELLPNVLKLVPGYKVKVIXKPMDFSTIREKLSSGQPNLETFALDVR 1934
Db 1331 TLLKEAMROECSPWFLQPVDSKEVDPDYDVIRKPNLRTMMNKIKORIYNKPIEVRNDFQ 1390
QY 1935 LVFDCNETENEDSDIGRAGHNRKY 1960
Db 1391 LLLSNCETNEPENIEYKLSRLHDF 1416
RESULT 31
Q8CAU9
ID Q8CAU9 PRELIMINARY; PRT; 657 AA.
AC Q8CAU9, (Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE Mus musculus 16 days neonate thymus cDNA, RIKEN full-length enriched
DE library, clone: A130044E01 product: bromodomain adjacent to zinc finger
DE domain, 1B, full insert sequence. (Fragment).
GN Name=Bazib;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Thymus;
RX MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning."
RL Meth. Enzymol. 303:19-44 (1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Thymus;
RX MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RA RIKEN FANTOM Consortium;
RT "Functional annotation of a full-length mouse cDNA collection."
RL Nature 409:685-690 (2001).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Thymus;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RL 60,770 full-length cDNAs."
RN Nature 420:563-573 (2002).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Thymus;
RX MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RA Carninci P., Shibata Y., Hayatsu M., Sugahara Y., Shibata K., Itoh M.,
RA Konno H., Okazaki Y., Muramatsu N., Hayashizaki Y.;
RT "Normalization and subtraction of cap-trapper-selected cDNAs to
RT prepare full-length cDNA libraries for rapid discovery of new genes."
RL Genome Res. 10:1617-1630 (2000).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Thymus;
RX MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RA Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M.,
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA Fujiwara S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
RL sequencing pipeline with 384 multicapillary sequencer."
RL Genome Res. 10:1757-1771 (2000).

1822	DSASTSSSLKRG	---	NK 1833
417	DDDEALSPKGRGPQVRLPIKTKGRFGPSPRSQRDPGRYPSRQOSTPKNTAKSASK	476	
1836	DLOKRG	---	1841
477	NLRKTRSAPTTETSLRVGSRSTRHSALQDVVELLSPHSKRGRKGADHTPEHSPSP	536	
1842	-----MEENTSINLSKQESPTSVKKPRDDSK	-----D 1869	
537	TNFRVSTSRSRQLPIPLNTAESLSQHSKSRGRKQSTESSVPVLRNRSSSGRQGVH	596	
1870	LALCSMILTEMETHEDAWPELLPVNLKLVGKYKVIKPMDFSTIREKLSSGQYPNLETF	1929	
597	LSAFEQLVVELVRHDDSNPFLKLVSKIQVPDYDIKPIALNIIREKVNKCYKLA	656	
1930	ALDVLVDFNCETNEEDSDIGRAGHNRKYP	1961	
657	IDDIELMFSCNCFEYNPRNTSEAKAGTRLOQAF	688	
RESULT 33			
Q7RN75	PRELIMINARY;		
ID Q7RN75	PRELIMINARY; PRT; 2095 AA.		
AC Q7RN75;			
DT 01-MAR-2004 (TremBLrel. 26, Created)			
DT 01-MAR-2004 (TremBLrel. 26, Last sequence update)			
DT 01-MAR-2004 (TremBLrel. 26, Last annotation update)			
DE Hypothetical protein.			
GN Name=PY01947;			
OS Plasmodium yoelii yoelii.			
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.			
OX NCBI_TaxID=73239;			
RN [1]			
RP SEQUENCE FROM N.A.			
RC STRAIN=17XNL;			
EX PubMed=12368865; DOI=10.1038/nature01099;			
RA Carlton J.M., Arguioli S.V., Suh B.B., Kooij T.W., Perrea M.,			
RA Silva J.C., Ermolaeva M.D., Allen J.E., Selengut J.D., Koo H.L.,			
RA Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L.,			
RA Shallom S.J., van Aken S.E., Riedmuller S.B., Feldblum T.V.,			
RA Cho J.K., Quackenbush J., Sedegah M., Shoaihi A., Cummings L.M.,			
RA Florens L., Yates F.R. III, Raine J.D., Sinden R.E., Harris M.A.,			
RA Cunningham D.A., Preiser P.R., Bergman L.W., Vaidya A.B.,			
RA van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R.,			
RA Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,			
RA Carucci D.J.;			
RT "Genome sequence and comparative analysis of the model rodent malaria			
RT parasite Plasmodium yoelii yoelii";			
RL Nature 419:512-519(2002).			
CC -!- CAUTION: The sequence shown here is derived from an			
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is			
CC preliminary data.			
DR EMBL; AABL01000531; EAA21352.1; --			
KW Hypothetical protein.			
SQ SEQUENCE 2095 AA; 241540 MW; 0500EC8A34D13677 CRC64;			
Query Match 4.0%; Score 409.5; DB 2; Length 2095;			
Best Local Similarity 18.7%; Pred. No. 2.5e-07;			
Matches 408; Conservative 326; Mismatches 703; Indels 741; Gaps 97;			
QY 2 GQTSTSSGGNRKCNQEQSKNQPLDARVDK	-----	32	
DB 250 GIKYKKGNNKNNKKKSNK	--EAKVDKKINISNNDKTEIDKYNLTNPELQNTQ	307	
QY 33	-----IKDKKP	-----RKKAMES-SGNSDSDSGTSSTSS	EGIS 65
DB 308 ESSISFQNIHVQDKCNKILNLCCKDKMSVNTSDQNSYKSLSLSCIKNEELS	367		
QY 66	SSSDDDLEED	-----EEDDOSIESEDDSDS	95
DB 368 ENKOTENLRKSNNNKSYLDSKBTITVDYTKLITPLMLSEHNTSI-KNEONKNDLEN	426		

1822	DSASTSSSLKRG	---	NK 1833
417	DDDEALSPKGRGPQVRLPIKTKGRFGPSPRSQRDPGRYPSRQOSTPKNTAKSASK	476	
1836	DLOKRG	---	1841
477	NLRKTRSAPTTETSLRVGSRSTRHSALQDVVELLSPHSKRGRKGADHTPEHSPSP	536	
1842	-----MEENTSINLSKQESPTSVKPKRDSK	-----D 1869	
537	TNFRVSTSRSRQLPIPLNTAESLSQHSKSRGRKQSTESSVPVLRNRSSGROGVH	596	
1870	LALCSMILTEMETHEDAWPELLPVNLKLVGKYKVIKPMDFSTIREKLSSGOYPNLETF	1929	
597	LSAFEQLVVELVRHDDSNPFLKLVSKIQVPDYDIKPIALNIIREKVNKCYKLA	656	
1930	ALDVLVDFNCETNEEDSDIGRAGHNRKYP	1961	
657	IDDIELMFSNCFEYNPRNTSEAKAGTRLOQAF	688	
RESULT 33			
Q7RN75	PRELIMINARY;		
ID	Q7RN75	PRELIMINARY;	PRT; 2095 AA.
AC	Q7RN75;		
DT	01-MAR-2004 (Tremblrel. 26, Created)		
DT	01-MAR-2004 (Tremblrel. 26, Last sequence update)		
DT	01-MAR-2004 (Tremblrel. 26, Last annotation update)		
DE	Hypothetical protein.		
GN	Name=PY01947;		
OS	Plasmodium yoelii yoelii.		
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.		
OX	NCBI_TaxID=73239;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=17XNL;		
EX	PubMed=12368865; DOI=10.1038/nature01099;		
RA	Carlton J.M., Arguioli S.V., Suh B.B., Kooij T.W., Perrea M.,		
RA	Silva J.C., Ermolaeva M.D., Allen J.E., Selengut J.D., Koo H.L.,		
RA	Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L.,		
RA	Shallom S.J., van Aken S.E., Riedmuller S.B., Feldblum T.V.,		
RA	Cho J.K., Quackenbush J., Sedegah M., Shoaihi A., Cummings L.M.,		
RA	Florens L., Yates F.R. III, Raine J.D., Sinden R.E., Harris M.A.,		
RA	Cunningham D.A., Preiser P.R., Bergman L.W., Vaidya A.B.,		
RA	van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R.,		
RA	Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,		
RA	Carucci D.J.;		
RT	"Genome sequence and comparative analysis of the model rodent malaria		
RT	parasite Plasmodium yoelii yoelii."		
RL	Nature 419:512-519(2002).		
CC	!- CAUTION: The sequence shown here is derived from an		
CC	EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is		
CC	preliminary data.		
DR	EMBL; AABL01000531; EAA21352.1; --		
KW	Hypothetical protein.		
SQ	SEQUENCE 2095 AA; 241540 MW; 0500EC8A34D13677 CRC64;		
Query Match	4.0%; Score 409.5; DB 2; Length 2095;		
Best Local Similarity	18.7%; Pred. No. 2.5e-07;		
Matches	408; Conservative 326; Mismatches 703; Indels 741; Gaps 97;		
2	QOTKSTSGGNRKCQEQSKNQPLDARVDK	-----	32
250	GIKYKKGNNKKNKKKQNK--EAKVDKKNISNENDKTEIDKYNLTYPFELQNTQ	307	
33	-----IKDKKP-----RKKAMES--SNSDSDSGTSSTSS-----EGIS 65		
308	ESSISFQNIHVQDKCNKILNLCCKDKMSVNTSDQNSYTSKYLSLSCIKNEELS	367	
66	SSSDDDLEED-----EEDDSISEDDSDSSES 95		
368	ENKOTENLRKSNKNSYLSKBTITVDYTKLITPLMLSEHNTSI-KNEONKNDLEN	426	
SEQUENCE FROM N.A.			
RC	STRATN=C57BL/6J; TISSUE=Cerebellum;		
RA	Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,		
RA	Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,		
RA	Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,		
RA	Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,		
RA	Katoh H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,		
RA	Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,		
RA	Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,		
RA	Saito D., Saibata K., Sakai K., Sakai K., Sakazume N., Sano H.,		
RA	Sasaki R., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,		
RA	Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,		
RA	Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.,		
RL	Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AK047427; BAC33055.1; --		
DR	HSSP; Q9UIG0; 1F62.		
DR	MGD; MGI:3039627; BC065123.		
DR	GO; GO:0000151; C-ubiquitin ligase complex; IEA.		
DR	GO; GO:0003677; F-DNA binding; IEA.		
DR	GO; GO:0004842; F-ubiquitin-protein ligase activity; IEA.		
DR	GO; GO:0008270; F-zinc ion binding; IEA.		
DR	GO; GO:0016567; P-protein ubiquitination; IEA.		
DR	GO; GO:0006355; P-regulation of transcription, DNA-dependent; IEA.		
DR	InterPro; IPR001487; Bromodomain		
DR	InterPro; IPR011011; FYVE PHD_Znf.		
DR	InterPro; IPR001965; Znf PHD.		
DR	InterPro; IPR001841; Znf ring.		
DR	Pfam; PF00439; Bromodomain; 1.		
DR	Pfam; PF00628; PHD; 1.		
DR	PRINTS; PR00503; BROMODOMAIN.		
DR	SMART; SM00297; BROMO; 1.		
DR	SMART; SM00249; PHD; 1.		
DR	SMART; SM00184; RING; 1.		
DR	PROSITE; PS00633; BROMODOMAIN 1; 1.		
DR	PROSITE; PS00014; BROMODOMAIN 2; 1.		
DR	PROSITE; PS01359; ZF_PHD_1; 1.		
DR	PROSITE; PS00		

RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
 RA Jamal M.B., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai X.,
 RA Laekko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
 RA Williams S.M., Woodgett J., Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
 RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*.";
 RT Science 287:2185-2195(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426065; PubMed=12537568;
 RA Celnik S.E., Wheeler D.A., Kronmiller B., Carlson J.W., Halpern A.,
 RA Patel S., Adams M., Champagne M., Dugan S.P., Frise E., Hodgson A.,
 RA George R.A., Hoskins R.A., Laverty T., Muzny D.M., Nelson C.R.,
 RA Pacleb J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,
 RA Svirskas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
 RA Weinstein G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.;
 RT "Finishing a whole-genome shotgun: Release 3 of the *Drosophila*
 RT *melanogaster* euchromatic genome sequence.";
 RT Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426070; PubMed=12537573;
 RA Kaminker J.S., Bergman C.M., Kronmiller B., Carlson J., Svirskas R.,
 RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,
 RA Ashburner M., Celnik S.E.;
 RT "The transposable elements of the *Drosophila melanogaster* euchromatin:
 RT a genomics perspective.";
 RT Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).
 RN [4]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426069; PubMed=12537572;
 RA Mira S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
 RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,
 RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
 RA Bettencourt B.R., Celnik S.E., de Grey A.D., Drysdale R.A.,
 RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
 RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
 RA Lewis S.E.;
 RT "Annotation of the *Drosophila melanogaster* euchromatic genome: a
 RT systematic review.";
 RT Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).
 RN [5]
 RP SEQUENCE FROM N.A.
 RG FlyBase.
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
 RN [6]
 RP SEQUENCE FROM N.A.
 RG FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AE003808; AM70936.1; --
 DR HSSP; P11362; LEVT.
 DR FlyBase; Fggn0013988; Stm-Mlck.
 DR GO; GO:0004685; F-actin- and calmodulin-dependent protein k...; IDA.
 DR GO; GO:0004687; F-actin-light-chain kinase activity; IDA.
 DR InterPro; IPR003961; FN_III.
 DR InterPro; IPR008957; FN_III-like.
 DR InterPro; IPR007110; IG-like.
 DR InterPro; IPR003598; IG_c2.

DR InterPro; IPR003006; Ig_MHC.
 DR InterPro; IPR011009; Kinase like.
 DR InterPro; IPR000719; Prot Kinase.
 DR InterPro; IPR002290; Ser Thr_kinase.
 DR InterPro; IPR008271; Ser_thr_kin_AS.
 DR Pfam; PF00041; fn3; 1.
 DR Pfam; PF00047; Ig; 5.
 DR Pfam; PF00069; Pkinase; 1.
 DR ProDom; PD000001; Prot_kinase; 1.
 DR SMART; SM00060; FN3_2.
 DR SMART; SM00408; IGC2; 12.
 DR SMART; SM00220; S_TKc; 1.
 DR PROSITE; PS50853; FN3; 2.
 DR PROSITE; PS50835; IG LIKE; 27.
 DR PROSITE; PS00290; IG_MHC; UNKNOWN_1.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; UNKNOWN_1.
 DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
 DR PROSITE; PS00108; PROTEIN_KINASE_ST; UNKNOWN_1.
 KW ATP-binding; Kinase; Transferase.
 SQ SEQUENCE 9270 AA; 1036760 MW; 0DD82D040B27C2AD CRC64;
 Query Match 4.0%; Score 402; DB 2; Length 9270;
 Best Local Similarity 18.6%; Pred. No. 2.6e-06;
 Matches 416; Conservative 388; Mismatches 903; Indels 534; Gaps 89;
 QY 21 SKNQPLDARVD-----KIKDKPKKXAMESSNSDS-----DSGTSS 57
 DB 2222 SPESALDITDIDKEGSGQEDIEAKPKMLESQQLDSYKQTETQEEVPKTDDETKK 2281
 QY 58 DTSSGSISSDS-----DLLEDEEEDDSQSESE-----DDSDSSEAHQKNNQVL 106
 DB 2282 DIEVESKLENQNELVAKDEQKADKVSQEKQESKQQTQTEVDTDQKSTEVVSKASPENI 2341
 QY 107 LHGISDPKADG-QKATEKAQK-----RIHQPLPLAFESQTHSFQ---- 145
 DB 2342 LEALSEKLSQSPNNATQNDIEITNTEQDILDNDINTEKSKSFKUREHIVHTFDGKP 2401
 QY 146 -SQKQPOVLSQQLPFIFQS-SQAKESVNVKHTSVIQTGLVSNVKNPLSLVNOAKKETVM 203
 DB 2402 PEEQTEKELVEKLTESLFESCPEATEHVITQYKEIK-TNIIITKAAIQLIDDSNLTGPK 2460
 QY 204 KLIVPS-----PDVLKAGNKVTS-----ESSLL-----TSELRSKKE 236
 DB 2461 SLLVPLKLVNLEKLSLTQTVLIDKSSKEMIGLQQLMDIFILDDLDERTEKINPKIE 2520
 QY 237 QYKQAFPSQ---LKKQESSKSLKVIKALSNPKATSSSP-----AHPKQTLNNHNPFF 287
 DB 2521 NIKKILLSEYDIEKKG-----QNTAVVNGKIKLITEKILDCIEETFKQITESQNKND 2575
 QY 288 LTNALLGNHQPNGVITQSVIQ--EAPLALTTTKMQSKINENIAAASSTPFSSPVNLSTSG 345
 DB 2576 AAGDIKKSETEDVVDHSIEKIEEPKRSKDKLDKFELEEKELKASA----- 2622
 QY 346 RRTPGNQTPVMPSPASPIHLSQKQEKAVSNV-NPVKTOHHSHPAKSLVEQFGTSDSDIPS 404
 DB 2623 -KKQGDQDIEQKQKQPEVSEVVAEKISEGKIEEPKPEEMDTAKS--EKATVLDKQVLE 2679
 QY 405 SKDSEDSNDEDEED-----EEDEDEDEDDSDSDSQSDSDSSESD 446
 DB 2680 EKELEASAEKQGDQDVEKKQKQPEVSEVVAEKISEETIEEPKPEVKQTEIKSEKATALD 2739
 QY 447 TEGSEEDDDKDDQDESDDTEGKTSMLKNTTSSVKSPSMLTGHSTPRNLHIAKAPG 506
 DB 2740 KQVLEKELEASAKQKQCDQV--EKKSQK--PEVSEIVAEXIS-----EKTIEEPKPE 2789
 QY 507 SAPAALCESQSPAPLGTSSSTLTSSPHSGTSKRRVTDRELRPLPLEYQWQRETRINF 566
 DB 2790 VKDTEIKSE-----KATALDKQVLEKELEA----- 2815
 QY 567 GGRLOGEVAYAPCGKKLQYQPEVTKYLSRNGIMDISRDNFSFSKAKIVGDFEYARDGPQ 626
 DB 2816 SAQKQGDQV-----EKKSQKQPEVSEVVAEK-----ISEETIEEPKPEVKETEVSSEKAT 2866

DR	Pfam; PF00041; fn3; 11.	
DR	Pfam; PF00047; ig; 6.	
DR	Pfam; PF00069; Pkinase; 1.	
DR	SMART; SM00060; FN3; 11.	
DR	SMART; SM00408; IGC2; 16.	
DR	SMART; SM00220; S_TKC; 1.	
DR	PROSITE; PS00853; FN3; 11.	
DR	PROSITE; PS00835; IG_LIKE; 38.	
DR	PROSITE; PS00107; PROTEIN KINASE ATP; 1.	
DR	PROSITE; PS00111; PROTEIN KINASE DOM; 1.	
DR	PROSITE; PS00108; PROTEIN KINASE ST; 1.	
DR	PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN 1.	
KW	ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.	
SQ	SEQUENCE 18519 AA; 2051739 MW; 6A8441C5D0BA7729 CRC64;	
Query Match 3.9%; Score 397; DB 2; Length 18519;		
Best Local Similarity 18.1%; Pred. No. 8.6e-06;		
Matches 427; Conservative 359; Mismatches 923; Indels 648; Gaps 80;		
QY	6 STSGGGNKKQEQSQKNQPLDARDVKK-----DKPKRKAM- 43	
DB	9279 STSDASEVAQVQOSEAQVTEKPEAKLESKMKMTDTIKESDNKETVDEKPKKKVLK 9338	
QY	44 ESSNSDSGTSSTDS---SEGISSDDDL-----EEDSEEDQIE-----ESE 87	
DB	9339 KXTEKSDTSETSETSAVESAGPSESTONVAADVDEKKQKQKQKEDKQKLEAIAIGKST 9398	
QY	88 DDDSDSESAQHNQNNQVLLHGSDPKAQKATEKAQEKRIHQPLAPFESQ-----140	
DB	9399 EQSKLEAEAKLKRAE-----EDAAKQKQKTEAAKAAAEKLEKQKQKQKQKAAEA 9452	
QY	141 -----THSPSQSQK---QPVLSSQQLPFIQSS-----QAKESVNHKTSVIQSTG 183	
DB	9453 DAVKQNELDEQNKLEATKKLAAEKLKLEQSAAKSKQAABEQAKLDAQTKAAAEKQTG 9512	
QY	184 LVSN-----VKPLSLVNOAKETVMKLI VSPDVLK-----AGNKT-----220	
DB	9513 LEKDEKSNKDSGNETVEBKPKKKVLKKTETKSDSSISQKSDTSKTVAESAGSSESTQK 9572	
QY	221 -----SEBSLLTSELRSKR---BOYQAPPSQL-----KKQESSKSLKVI 259	
DB	9573 VADATSKQKQETKKQKLEAIIYAKSADEKSKLETESKLIKAAEDAAKKQKQKEDKLE 9632	
QY	260 AALSNPKATSSGPAHPKOTLENNHPNPLTNALLGNHPNGVQSVIOEAPLA-----312	
DB	9633 ADVASKKA-----AAEKLEK-----QAQIKKAAEADAVKKQK 9666	
QY	313 -LTTTKTQMSKINENTAAASSTPFSPVNLSTSGRTPGNQTPVM-----PSASP 361	
DB	9667 ELAEKQKLESEATKKAAAEKLEBQQAQINKAAEADAVKKQKLEKKNLEANKKSAAE 9726	
QY	362 ILHSQKKEKAVNNVNPVKTQHHSHHPAKSLVQFRTGTDSDIPSSKQSDSDNEDEEDDE 421	
DB	9727 KULEEESAASKQTVBQKDLAQTKETAKQKQKLEKDDKSDTSKESKETTVEKPKKK 9786	
QY	422 EDEEDDEDDSDSQS-----ESDSNSEDTEGSEEDDDDDKQDSDSDSDTEGEKTS 473	
DB	9787 VLKKTETKSDSSTISQKSVTSKTVESGGPSESETQKVADAARKQKQKQKQKQKQKQK 9845	
QY	474 MKLNTTSSVKSPSMLTGHSTPRNLHIAKPGSAPALCSQSQSPAFLTSSSTLTSSP 533	
DB	9846 KKSADKSKLEAS-----KLKAAAEVAEAAKQKQKQKQKQKQKQKQKQKQKQKQK 9888	
QY	534 HSGTSKRRRTVERELRIPLYEGWQRETRIRNFGGRLOGEVAYAPCGKKLQYQPEVIKY 593	
DB	9889 -----KKAAEKLEL-----EKQAQKK-----AAEADAVKKEKLEAKQKQL 9925	
QY	594 LSRNGIMDISRNFSFSAKIRVGDGYEARDGQEQMOWCLLKBEDVTPRIRAMEGRRGRPP 653	
DB	9926 ESEAAKKAEEKKLEEQ-KKKDAETASIEKQKQKQKQKQKQKQKQKQKQKQKQKQKQK 9976	
QY	654 NPDRQARRESRRRRKGRPPNVGNAEFLDNADAKLLRLKLOAQETARQAQKILKRLKQK 713	

DB	9977 SAEKOKLSESTKSKTTEAPK-----ESVDEKPKPKVKLKKKTEKSDSSIS-----OK 10023	
QY	714 QEQARVAKEAKQ-----QAIMAAEBEKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQ 768	
DB	10024 SDTAKTVAESAGQSDSETQKVSEADKAHQKE-----SDEKQK-----LSEIAAKKS 10071	
QY	769 LEAKKKKKEEAAANAKLLBAE--KRIKEKEMRRQQAALLKHQERRRRRRHMLMKAMEARK 826	
DB	10072 ARQKSKLETEAKTKKVIEDESAKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQK 10127	
QY	827 KAEERLERKQEKDEKRLNKKERK---LEQRELELE-----MAKEL-----863	
DB	10128 KKPTSEKQ 10187	
QY	864 ----KKPNEDMCLADQKPLPELPRIPLGLVLSG-----TFSDCLMNVVQ 902	
DB	10188 DSKTKEPREKKKIIKKKQDXTTKPOEASKELSDSDRIDLESIDLSLDIVTESDDDLSTAS 10247	
QY	903 F--LRFNGKVLGFDVNI-----DVPNLVLQEGLLNIGDSMGVEQDILLVRLLSNAVCD 953	
DB	10248 TIKLQKESDESIGDSRMGQTSSEADSPFIS--QPVSATVTMAGEAK-FTVKFSRKPIYV 10304	
QY	954 PGLITGYAKAKTALGEH-----LLNVGVNRDNVSEILOIF-----MEAH 991	
DB	10305 KWMRRDRIRVAYGKASVETTTDDSVLVTKNIDGKDVGNIVYAVFQSEYRSAMARLDLRVP 10364	
QY	992 CQOTELTESLTKTAPQAHTPAQKASVLAFLI-----NELACSVSVSEIDK 1037	
DB	10365 CKIT-----LESSNAPEIVAGKNLDLSFKISGYPLPTNIELHNNENLRTSDVDFDD 10419	
QY	1038 NIDYMSNLRDRKVVVEGKLR-----KLRI-----IHAKTKGKRDTS-----1073	
DB	10420 SISII--RMKRLKLEDSGEIKIIGKNDSSSEDQLRIPINVIEWTSKPTSLQVTSSTERETVTL 10477	
QY	1074 -----GGIDGEEQHPGTPTPCRKR-----KGG 1099	
DB	10478 TWSLPTELNGSVNYLVERKTVDCGRWRHACTVTTDSRAVDGLFSGTGYVFRVAVNGA 10537	
QY	1100 DSDYDDDDDDDDDDQDQDDE-----DEEDKQKQKKTIDICEDD-----1140	
DB	10538 GQASPSDTIEATQAEIEIDEIVPTSPVEKVEPKPKPKPKPKPKPKPKPKPKPKPKPKPK 10597	
QY	1141 -----EGDQAASVEELE-----KQIEKLSQKQSYRRKLPDASHS 1175	
DB	10598 ENNLGKSGKDEPATSGESGTSNQNEESAQLNTSFTSTEQHGTQKQVRKGTKSLTRSLN 10657	
QY	1176 LRSVFGPDYRRRYWILLPRCGGI FVEGMESEGELEETAKEREKLUKKAESVQIKKEPMFT 1235	
DB	10658 IRESIDIDAD-----VVEVEYDEQDGDIPSDPTTSGTYAFDKIEEAPART 10701	
QY	1236 SGDSLNCNTDHCQEKDLKEKDNLTNLFQKPG-----SFSKL 1273	
DB	10702 SGEMA-----MAEKDSDAMEVRGLNKLKSKGKGEGTSTKSSSKTKQKESALSVOEM 10755	
QY	1274 SKLLEVAKMPPPESEVMTPKPNAGANGCTLSYQNSGKHSIGSVQSTATQSNVSKAOSNNUL 1333	
DB	10756 NKSLKKKKEGAEATAASDFIENADQTGMSIQDLNKMKKKVESGEATQINDASNNK-- 10813	
QY	1334 NTGSSGPKGFYSPLPNDQLLTKLTETKNQWFSLLPRT-PCDDTSLTHADMSTASLVTPOS 1392	
DB	10814 -----DAELSIQDSQOSLKKKSESVTSGQDQKQSEVEDDKMTIOSLKKSIKKKPPS 10867	
QY	1393 QPPSKSPPTAPLPGSSAQNVPVGLN-----PFALSPLOQKGVSN-----MGLQFCGW 1440	
DB	10868 REVSGGKSEKSEKSEKSEKSEKSEKSEKSEKSEKSEKSEKSEKSEKSEKSEKSEKSEKSE 10923	
QY	1441 PTGVVTSNIPFTLSVPSLGSGLGSEGNNSFLTNSVASSKSESPPVQNEKATSQAAPAAV 1500	
DB	10924 ---ALNKMKKKGDVDGVEGNINIGKSDQDLSVNDIDAEALSTSEVEN-----ASQ 10972	
QY	1501 EVAKVPDPPSPKPIPEEMQFGHWRIIDPEDLKALLKVLHLRIGIRKALQKQKQKQKQKQK 1560	
DB	10973 NLGATADSDG-----DSLSLQTLKRISSKKGIHGEASK-----11006	

QY 1561 QACLKXKQVAILIENENENOVTRIDVENWSVEBOAMEMDLSVLQOVVEDLERRVASASIQ 1620
Db 11007 -----LGEKSGSDSFTLQDLYEELKAKEDAVEAG-----AETSNADOSEAKETSLE 11052
QY 1621 V-----KGWMCPEFASEREDLVVFEHKSFTK-----LCKEHDGEFTGEDSSAHALERKSD 1671
Db 11053 VRDMKKMKKKQVSGTAENLIGESNRDETSMEIRDINTQHSNQ--TGEDSSFTNFQKQDO 11111
QY 1672 NPLDIATRLADLERNIERIBEDTAPGLRVWRRLSEARSAAQVALCICQI--QKSIWAEK 1731
Db 11112 EQYSNV--MKOVSKKLARONAEELQSG-----KLIPPTNEEKTGLALTGNKNL---- 11158
QY 1732 STMKYVCQICRKGDNELLLLCDGCDKGCHTYCHRPKTIIPDGWFCPACITAKASGQTL 1791
Db 11159 -----KKGENE-----KTKFEAKHLGSSASDSLSAESTLRSK 11191
QY 1792 KIKLHV-KGKTNESKKGKVTLTGDTEDSDASTSSSLRGKNDLQKRKMEENTSINL 1850
Db 11192 KTKKGEVKSLSIDMKQDKTTLATLLEDLAKTTSAESEAEHLVALQNKETSLAM 11251
QY 1851 -----SKOESFTSVKPKRDSKDLALCSMLTETMETHEADWPFLLPVNLKLVPG 1900
Db 11252 RKRVSFDSSTKSEGTEDVIPDKNRDSDKMSITGIKKMSQKSESA-----EAQKNESPE 11306
QY 1901 YKKVIKPNWDFSTIREKLSS-----GQYPNL--ETFALDVLVFDNCETFNEDSDIGR 1952
Db 11307 VKEI--SSPEETKLSKKKSKADRQNGTEANLGDITDKOYLSVTDKQSLKSEAS-GQ 11363
QY 1953 AGHNRKPYEKKWDTTF 1969
Db 11364 AEKSIKAPNKSIVTTSF 11380
RESULT 37
Q8ISF7
ID Q8ISF7 PRELIMINARY; PRT; 18534 AA.
AC Q8ISF7;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DE 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE 2MDA.1 protein.
GN Name=isoF;
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Feloderinae; Caenorhabditis.
OC NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22269627; PubMed=12381307; DOI=10.1016/S0022-2836(02)00970-1;
RA Flaherty D., Gernert K., Shmeleva N., Tang X., Mercer K.,
RA Borodovsky M., Benian G.;
RT "Ttins in Caenorhabditis elegans with Unusual Features: Coiled-coil
RT Domains, Novel Regulation of Kinase Activity and Two New Possible
RT Elastic Regions."
RL J. Mol. Biol. 323:533-549(2002).
CC -I- SIMILARITY: Belongs to the Ser/Thr protein kinase family.
DR EMBL; AY130758; AAN61517.1; -.
DR PIR; E89066; E89066.
DR PIR; T33247; T33247.
DR HSSP; Q10466; 1BPV.
DR GO; GO:0005524; F-ATP binding; IEA.
DR GO; GO:0004674; F:protein serine/threonine kinase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0006468; P:protein amino acid phosphorylation; IEA.
DR InterPro; IPR008938; ARM.
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR008957; FN_III-like.
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003598; IG_c2.
DR InterPro; IPR011009; Kinase like.
DR InterPro; IPR006025; Pept M.Zn.BS.
DR InterPro; IPR000719; Prot_kinase.

DR InterPro; IPR002290; Ser_thr_kinase.
DR InterPro; IPR008271; Ser_thr_kin_AS.
DR Pfam; PF00041; fn3; 11.
DR Pfam; PF00047; ig; 6.
DR Pfam; PF00069; Pkinase; 1.
DR SMART; SM00060; FN3; 11.
DR SMART; SM00408; Igc2; 16.
DR SMART; SM00220; S_TKc; 1.
DR PROSITE; PS50853; FN3; 11.
DR PROSITE; PS50835; IG_LIKE; 38.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN 1.
KW ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.
SQ SEQUENCE 18534 AA; 2053374 MW; BC8A682B943C8COA CRC64;
Query Match 3.9%; Score 397; DB 2; Length 18534;
Best Local Similarity 18.1%; Pred. No. 8,6e-06;
Matches 427; Conservative 359; Mismatches 923; Indels 648; Gaps 80;
QY 6 STSGGGRKCNQEQSKNQPLDARVDYKIK-----DKKPRKAM- 43
Db 9279 STSDASEVQVQOSEAEQVTEKPTAKLESKSKWTEDTTKESDNKVTDEKPKKVLK 9338
QY 44 ESSNSDSGTSDDTS---SEGISSSSDDL-----EDEEEDQSIE-----ESE 87
Db 9339 KTKESDSTISSETSETSAVESAGPSSESTQNVAAVDKKEKQKQETDEKQKLEAETAGKST 9398
QY 88 DDDSDSEAEKHKSNNQVLLHGIDPKADGQKATEKAOEKRIHQPLPLAFESQ----- 140
Db 9399 EQSKLEAEAKLRAAE-----EDAANKQEKTEAASKAAAEKLEKQAINKAAEA 9452
QY 141 -----THSFQSQK---QPQVLSQOLPFIFOSS-----QAKESVNKHTSVQSTG 183
Db 9453 DAVKQNELDQNKLEATKLAAEKLEEQSAKSAKQAEQAKLQATRAKAAEKQTG 9512
QY 184 LVSN-----VKPLSLVNOAKKETYMKLIVPSDVLK-----AGNKNV----- 220
Db 9513 LEKDEKSNKSGSNSTVEEKPKKYLKKTKEKSSISQKSDTSKTVAESGSSSESTQK 9572
QY 221 -----SEESLLTSELRSKR---EQYQAPPSQ-----KKQESSKSLKKVI 259
Db 9573 VADATSKQKQETDKKQLEAEITAKSADEKSKLETESKLIKAAEDAAKQKQEKEDKLLE 9632
QY 260 AALSNPKATSSPAHPKQTLNHNPNPFLTNALLGNHOPNGVQISVIOEAPLA----- 312
Db 9633 ADVASKKA-----AAEKLELEK-----QAQIKKAAEADAVKQK 9666
QY 313 -LTTTKMQSKINENIARAASSTPFSSPVNLSTSGRTPGNQTPVM-----PSASP 361
Db 9667 ELAEKQKLESEAAATKKAEEKLEEQAINKAAEADAVKQKQELDERKNKLEANKKAAE 9726
QY 362 ILHSGKGEKAVSNVNVNPKVQTHSHPAKSLVQEPFGTSDIPSSKSDSDSDEDEEDDEE 421
Db 9727 KKLUEESAASKQTVQEAQKLDQATKEKTAETKQGLEKDDKSTKDSKSETVDEKPKKK 9786
QY 422 EDEEDEDDESDDSQS-----ESDSNSESDETEGSEEDDDDKQDESOSDTEGEKTS 473
Db 9787 VLKKKTEKSDSISQKSVTSKTVVSEGGPSESTQKVADAARKQKETDEKQK-LEAEITA 9845
QY 474 MKLNKTTSSVKSPNSLTHGSTPNRLHIKAPGSAPALCSSESQSPALGTSSSLTSSP 533
Db 9846 KGADEKSKLEAES-----LUKKAEEVAEAAKKQKQEKDQLKLDTEAAS----- 9888
QY 534 HSGTSKRRRVTDRELRIPLEYGMORETRIRNFGRLQGEVAYVAPCGKQLRQYEVIKY 593
Db 9889 -----KKAEEKLEL-----EKQAQIKK-----AAEADAVKKEKLEAKQKL 9925
QY 594 LSRNGIMDISRDNFSFSAKIRVGDPFYEARDGQENQWCLLKEEDVPIPRAMEGRGRPP 653
Db 9926 ESEAATKKAEEKLEKLEEQ-KKKOAEATASIEKQEQEKLQEQSKL-----EVDAKK----- 9976

Db 1173 -YMLEKKKKRMSLQNKMISSNSPEHVQFLPNEVMVDGTTIKNGDDDDDDDDDDKSVGKI 1231
Qy 186 SNVKPL-----SLVNAQ---KETYMKLIVSPDVLKAGNKNTSSESSILTSLSKREQ 237
Db 1232 SEIRSEFSYESFISEQKKIKKINLIKQNVVRNKSIIIOHNIKEDIKENVVG 1291
Qy 238 YKQAFPSOLKQESSKSLKVTAALSNPKATSSSPAHPKQTLNHNHPFLTNALLG--- 294
Db 1292 KHEDDEKVKYEDKE--KDDI-----NSKQONS-----QTSFNNDN--IINNIFGIN 1338
Qy 295 NH-QPQNGVI-----QSVIQEAPLALTTKMQSKINENIAAASFPFSSPNLSTGRR 347
Db 1339 NYVQENGVLDPFNSSNLIIKEENKGEKKEKVEKINED-----E 1378
Qy 348 TPGNQTVPNPSAPLHSGQEKAV--SNVNPVKTQHHSPAKSLVE-----QFGTSD 401
Db 1379 NDNNDENILKQD---IERQKKELELELEIKKKKELELELOLMENKQFIGNKLI 1435
Qy 402 IPSKXSDSDNDEDEDEDEDESD--SQSESNSSESDTE---GSEERDD 455
Db 1436 DDLQKDINCTEKKKKKKEAEEETKINIDQMVNEKEMDVNSKEREIVQVHNEIKNT 1495
Qy 456 DKQODESDSDTEGKTSW-----KLNKTTSSVKSPS-----MSLTGHSTPR 497
Db 1496 NNKEE-----EGKKNLKKEKEINDCLINDYINKQKKKKKKNWAMYGRPIVKRONR 1548
Qy 498 NLHI-----AKAGSAPALCSQSAPFLGTSS-----STLTSSPHSGTSKRRR 542
Db 1549 NINIKNDLKLYSSKSGFNDYAFYAEFRFEVITGYNSEPDYLDIDQAKNEENKNDI 1608
Qy 543 VDERELRIPLE-----YG---WORETRIRFGRLQGEVAYVAPCGKK--- 583
Db 1609 IHNNIIKISKOMENIVENSPHYGRPIVEKSKGNPNYKIK-STHNAILKKRKK 1667
Qy 584 -----LROYPEV-----IKLSRNGIMDISDNFSA-----KI 613
Db 1668 TLKKSISINSPFKMSSNNKIWKRTSINKNTIDNYYNSTIKKIHKQEVNEDQGYDLKT 1727
Qy 614 RVGDYEADG--POEQWCLLKEDVPIPRAMEG-----RRGPP 653
Db 1728 RKLYDALDERINQTOQKNLKITENITQVGKKHQNVSNIKNTGAMLIEKITRKGND 1787
Qy 654 NPDORAREESMRERK--GRPPVGNAEFLDNADAKLLRLQAO----- 696
Db 1788 NEEDQFSELKALEKLKELKELRTEELTEEEKKLEELNSLKVTEEEKKKKKWMEEM 1847
Qy 697 -----ETARQAAQIKLRLKQEQARVAKEAKKQQAIMAAE-----EKRKQBOI 742
Db 1848 KQMEEMKTELQKKKEEKLQYSRRQEQIRKLESLKKKEELKLEEMNKLEEQKREEI 1907
Qy 743 KIM---KQOEKIKRIQOIRMEKELRAQILEAKKKKKKEAANAKLLBAEKRIKEMPRQ 799
Db 1908 KQMEEEKEREKLIKKEQKKKEEMKQE-----EQRKQE---IKLMEEMRKPEE---QR 1956
Qy 800 QAVLLKHQERERRROHMLM---KAMEARKAAEKERLKQE--KRDE-----KRLN 845
Db 1957 KVELKWLVEEKKEBELKLEEQKQREERKQWEEEKKREEEKKEEIKQWEEEKIKQE 2016
Qy 846 KERKL-----EORLELEMAK--ELKKNEDMCLADQKPLPRLPGLVLSGT 893
Db 2017 BEKKKEEEKKQWEEEKKQWEEEKKKKEEKKEEKKEEKKEEKKEEKKEEKKEEKKEEK 2068
Qy 894 FSDCLMVQFLNFGKVLGFDVNDVNLVULQEGLLNIGDSMEVQDILLVLLSAAVCD 953
Db 2069 -----EIKFLVE----- 2076
Qy 954 PGLITGYKAKTALGHLNNGVNRDNVSE--ILQIFMEAHCGQTELTESLK-----TKAF 1006
Db 2077 -----MKEK-----EHIKYIENKENTSEKDLHNLFLEKIYNVNFSLYIKNDKESTQIL 2126
Qy 1007 QAHTPAQKASVLAFILNACSKSVVSEIDKNID-----YMSNLRDRKQWVBEGLRK 1058
Db 2127 KSDYEKENEDKKYLIQKNIS---VEBIEKEDNEDDNNNNNNNNKDDNVICBKKK 2183

Qy 1059 ----LRIIH-----AKTKGRDTSGGID-----LGEEOHPLGT 1087
Db 2184 NIFMTKISHMNHYDIFYQMEKQRIKNKPKRYRNTPNNIGNKGLLKNAVTNNSHIIRN 2243
Qy 1088 PTPGRKRR-----RKGGSDYDDDDDD-----SDD-----QGEDEDEDEE 1123
Db 2244 IINKKEEYENNMKMSSEDEKNGDMSDDNMQSDDMANDNTQSDDMANDNTQSDMANDNT 2303
Qy 1124 DKBOQKGGKT---DICDEDEGQAASVEELEKQ-IEKL-----SKQSQYRRKLFDASHL 1176
Db 2304 PNDDMANDNTPNDDMANDNTSDDMANDNIIITKERIEKIIIDMKAREBKKKLL---QM 2359
Qy 1177 RSVMPGPDYRRYRYYILLPRCGGIFVEGMESEGELEEIAKEREKLEKKAESVOIKEMEETS 1236
Db 2360 RNKMLKYNRVNKENII-----NMVE-----KKLONKNSDKDEIDDDLPDEEEVEIFTE 2408
Qy 1237 GDSL-----NCSNTHCEQKEDLKEKDNLTNLFQKPGFSFKLSKLLLEVAKMPES 1286
Db 2409 SEESFLKIDDKGPGNFKXTHIYRESPHMKGNMMEYFNKYESSNEDNKSYS-NKSDREN 2467
Qy 1287 EVMTPKPNAGANGCTLSYQNSGKHSLSVQSTQTSQSVNEKADSNLNFVTS 1337
Db 2468 DMYSNNNNMNSHNILYSNSNK-----LTSYSNLTGNKNNLYNQNS 2510

RESULT 44
GOBI HUMAN
ID GOBI_HUMAN STANDARD; PRT; 3259 AA.
AC Q14789; Q14398;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Golgi autoantigen, golgin subfamily B member 1 (Giantin) (Macrogolgin)
DE (Golgi complex-associated protein, 372-kDa) (GCP372).
GN Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94187728; PubMed=7511208;
RA Seelig H.P., Schranz P., Schroeter H., Wiemann C., Griffiths G.,
RA Renz M.;
RT "Molecular genetic analyses of a 376-kilodalton Golgi complex membrane
RT protein (giantin).";
RL Mol. Cell. Biol. 14:2564-2576 (1994).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94257116; PubMed=8198703; DOI=10.1006/jaut.1994.1006;
RA Seelig H.P., Schranz P., Schroeter H., Wiemann C., Griffiths G.,
RA Renz M.;
RT "Macrogolgin -- a new 376 kD Golgi complex outer membrane protein as
RT target of antibodies in patients with rheumatic diseases and HIV
RT infections.";
RL J. Autoimmun. 7:67-91 (1994).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=95100974; PubMed=7802676;
RA Solda M., Miumi Y., Fujiwara T., Nishioka M., Ikehara Y.;
RT "Molecular cloning and sequence analysis of a human 372-kDa protein
RT localized in the Golgi complex.";
RL Biochem. Biophys. Res. Commun. 205:1399-1408 (1994).
CC -!- FUNCTION: May participate in forming intercisternal cross-bridges
CC of the Golgi complex.
CC -!- SUBUNIT: Homodimer; disulfide-linked.
CC -!- SUBCELLULAR LOCATION: Golgi; membrane-associated.
CC -!- DISEASE: Antigen in chronic rheumatoid arthritis and in the
CC autoimmune disease Sjogren's syndrome.
CC -!- SIMILARITY: Belongs to the golgin family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration


```

RESULT 46
FUTS_DROME
ID FUTS_DROME STANDARD; PRT; 5412 AA.
AC Q9W596; 076891;
DT 25-OCT-2004 (Rel. 45, Created)
DT 25-JAN-2005 (Rel. 46, Last sequence update)
DE Microtubule-associated protein futsch.
GN Names=futsch; ORFNames=CG3064;
OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A., FUNCTION, SUBCELLULAR LOCATION, TISSUE
RP SPECIFICITY, AND DEVELOPMENTAL STAGE.
RC STRAIN=Oregon-R;
RX PubMed=10839355; DOI=10.1016/S0896-6273(00)81169-1;
RA Hummel T., Krukkert K., Roos J., Davis G., Klammt C.;
RT "Drosophila futsch/22C10 is a MAP1B-like protein required for
RT dendritic and axonal development.";
RL Neuron 26:357-370(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkeley;
RX MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;
RA Adams M.D., Celnikier S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers J.-H.C., Blazef R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter A.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glöck A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
RA Hoslin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclet J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-P., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
RN [3]
RP GENOME REANNOTATION.
RX MEDLINE=22426069; PubMed=12537572;
RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochownik S.E.,
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
RA Bettencourt B.R., Celnikier S.E., de Grey A.D.N.J., Drysdale R.A.,
Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
Lewis S.E.;
RT "Annotation of the Drosophila melanogaster euchromatic genome: a
RT systematic review.";
RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).
RN [4]
RP SEQUENCE FROM N.A.
RX STRAIN=Oregon-R;
RX MEDLINE=20196011; PubMed=10731137; DOI=10.1126/science.287.5461.2220;
RA Benos P.V., Gatt M.K., Ashburner M., Murphy L., Harris D.,
RA Barrell P.G., Ferraz C., Vidal S., Brun C., Demailles J., Cadieu E.,
RA Dreano S., Gloux S., Lelaure V., Mottier S., Galibert F., Borkova D.,
RA Minana B., Kafatos F.C., Louis C., Siden-Kiamos I., Bolshakov S.,
RA Papagiannakis G., Spanos L., Cox S., Madueno E., de Pablos B.,
RA Modolell J., Peter A., Schoettler P., Werner M., Mourikioti F.,
RA Beinert N., Dowe G., Schaefer U., Jaekle H., Bucheton A.,
RA Callister D.M., Campbell L.A., Darlamitsou A., Henderson N.S.,
RA McMillan P.J., Salles C., Tait E.A., Valenti P., Saunders R.D.C.,
RA Glover D.M.;
RT "From sequence to chromosome: the tip of the X chromosome of D.
RT melanogaster.";
RL Science 287:2220-2222(2000).
RN [5]
RP FUNCTION, SUBCELLULAR LOCATION, AND TISSUE SPECIFICITY.
RX PubMed=10839356; DOI=10.1016/S0896-6273(00)81170-8;
RA Roos J., Hummel T., Ng N., Klammt C., Davis G.W.;
RT "Drosophila futsch regulates synaptic microtubule organization and is
RT necessary for synaptic growth.";
RL Neuron 26:371-382(2000).
RN [6]
RP FUNCTION, AND INTERACTION WITH FMR1.
RX PubMed=11733059; DOI=10.1016/S0092-8674(01)00589-X;
RA Zhang Y.Q., Bailey A.M., Matthies H.J.G., Renden R.B., Smith M.A.,
RA Spese S.D., Rubin G.M., Broadie K.S.;
RT "Drosophila fragile X-related gene regulates the MAP1B homolog futsch
RT to control synaptic structure and function.";
RL Cell 107:591-603(2001).
CC -!- FUNCTION: During embryogenesis, necessary for dendritic and axonal
CC organization and growth at the neuromuscular junction through the
CC regulation of the synaptic microtubule cytoskeleton. Microtubule
CC hairpin loops are found within a small subset of synaptic boutons
CC at the neuromuscular synapse, these loops are stabilized by
CC futsch. Loop morphology and dynamics suggest that rearrangement of
CC these microtubule-based loops is a critical component of the
CC process of bouton division and for subsequent nerve-terminal
CC growth and branching. Translation is repressed by Fmr1.
CC -!- SUBUNIT: Interacts with Fmr1.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic. Microtubule-associated.
CC -!- TISSUE SPECIFICITY: Neuronal cells within the PNS and CNS.
CC -!- DEVELOPMENTAL STAGE: All stages.
CC -!- MISCELLANEOUS: 'futsch' means 'gone' in German.
CC -!- CAUTION: Ref.4 sequence differs from that shown due to erroneous
CC gene model prediction.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE003420; AAF45622.3; --
CC EMBL; AL031128; CAA20006.1; ALT_SEQ.
CC PIR; T13564; T13564.
CC FlyBase; FBgn0015390; futsch.
CC GO; GO:0005875; C:microtubule associated complex; IDA.
CC GO; GO:0008017; F:microtubule binding; IDA.
CC GO; GO:0007409; P:axonogenesis; IMP.
CC GO; GO:0016359; P:dendrite morphogenesis; IMP.
CC GO; GO:0007456; P:eye morphogenesis (sensu Drosophila); NAS.
CC GO; GO:0000226; P:microtubule cytoskeleton organization and b. .; IMP.

```



```
QY 1129 KGKTTDICEDEGDOAAVSELEKQIEKLSKQOYRRKLPDASHLSRVNFGPDYRR 1188
DB 2511 KSVRSPLVASDH-EAAVAIED-----DAKSSISP-----DKKSR- 2544
QY 1189 RYWILLPROGIFVEGMEGEGELEIAKEREKLLKAESVOIKEEMPETSGDSLNCNTHC 1248
DB 2545 -----PGFVNETVSS--PIEATMEFSKIEVVEKSSIALSLQGGSGKK--QTD-- 2589
QY 1249 EOKEDLKEKNTNLFLOKPGFSKLSKLEVAKMPPESEVMTPKPNAGANGCTLSYQNSG 1308
DB 2590 SSPVDVAEGDFSHAVASVTPTLTTPAELAQIGAIAKTVSSPLDEA----- 2636
QY 1309 KSLSGSVQSTATQSVNEKADSNLNTGSSGKGFYSP--LPNDQLLKLITKRN-QWFS 1365
DB 2637 -----LRTPSAPEHISRADS-----PAECASEBIASQDKSPQVLKESRPAMWA 2680
QY 1366 LLPRTPCDTSILTHADMTASLVTPQSPKSPSPPTAPLGSAGNPNVGLNP----- 1418
DB 2681 -----ESKDDAA-----QLKSSVEDLRSVASTETSRPASGETASSPIEAPKDFABPE 2730
QY 1419 ----FALSPL--QVKGGSVMGLQFCGWPTGVVTSNIPFTLSVPSLGSGLGSEGNNSFL 1473
DB 2731 OAEKAVLPLTIELKGNLPLSS-----PVDVAHGDPFQPTSTTSPTVASVQPAELSKVD 2785
QY 1474 TSNVASS-----KSESPVQNEK-----ATSAQPAEVAEAKVDPFPSP----- 1511
DB 2786 IEKTASSPIDEAPKSLIGCPAERPEPSPAESAADAESVEKSKDASRPSPSVVESTKADST 2845
QY 1512 ----KPIPEMFGWRIIDPDRLKALLVHLRGKALQKQKHLDYITQACLK-- 1565
DB 2846 KGDISPPSPSVLEG-----PKD-----DVEKSSRRPSPVSAS-----ITGSTKDV 2888
QY 1566 NKDVAIIELENEENQV---TRDIVENWSEBQAMEMDLVQQVEDLERRVASASLQVK 1622
DB 2889 SRPASVWESVKDEHDKAESRESRIAKVESIDEACKSDSKSSQSDQKDEKSTLAS---- 2944
QY 1623 GWCPEPASEREDLVYFEHKSSTKCKEHDGFTGEDES---SAVALERKSNPDIADVT 1679
DB 2945 ----KEASRRSESV-----ESSKODAEKSESRPESVIAAGBPVPRESKPLDSKDT 2991
QY 1680 R-----LADLERNTERRIEEDIAPGLRVWRRLSEARSAQAQVLCIQOL----QKS 1726
DB 2992 SRPGSMVESVTADEKSEQQSRRESVAESVKADTKDGKQSEARSPSSVDELLKDDDBKQ 3051
QY 1727 IAWKSIIMKYVQICIRKGNBELLCLDCGCHTYCHRPKITIPDGMWPCACIACA 1786
DB 3052 ESRRQSITGSHKAMTWGDESPM---DKADKSEK--SRP-----ES 3088
QY 1787 SQGTJIKIKLHVKKTNESKGGKVTLTGDTDEDSASTSSSLKRGKND-LQKRMEN 1845
DB 3089 VAESIK-----HENTKDEESPL--GSRSDSVAESIKSDITKGEKSPLPSEVSRP 3136
QY 1846 TSI-----NLSKQESFTSVKKPKRDDSKD 1869
DB 3137 ESVGSGIKDEKASRRRESVAESVKP--ESSKD 3166
```

RESULT 47

```
Q7R6D1
ID Q7R6D1 PRELIMINARY; PRT; 1811 AA.
AC Q7R6D1;
DT 01-MAR-2004 (Tremblrel. 26, Created)
DT 01-MAR-2004 (Tremblrel. 26, Last sequence update)
DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)
DE GLP_574_37298_42733.
OS Giardia lamblia ATCC 50803.
OC Eukaryota; Diplomonadida; Hexamitidae; Giardiinae; Giardia.
OX NCBI TaxID=184922;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WB C6;
RA Morrison H.G., McArthur A.G., Adam R.D., Aley S.B., Gillin F.D.,
RA Olsen G.J., Sogin M.L.;
```

```
RT "Draft sequence of the Giardia lamblia genome.";
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
CC !- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AACB01000002; EAA42814.1; -.
DR GO; GO:0007242; P:intracellular signaling cascade; IEA.
DR InterPro; IPR001683; PX.
DR Pfam; PF00787; PX; 1.
DR PROSITE; PS0195; PX; 1.
SQ SEQUENCE 1811 AA; 201006 MW; DE7C286E89FDCDD6 CRC64;
```

Query Match 3.7%; Score 373; DB 2; Length 1811;
Best Local Similarity 20.7%; Pred. No. 5.1e-06;
Matches 305; Conservative 248; Mismatches 536; Indels 383; Gaps 63;

```
QY 26 LDARVDKTKKKPR-KKAWESSNS-----DSDSGTSDTSSSEGISSSSDSDDLEDEEE 78
DB 423 LDAQGEPAQADQPDIDQVAQVSADSMFGFDLPADASATVEPSTTGLAT-----BKP 473
QY 79 EDQSEIEESDDSDSEBAQHKNQVLLHGISDPKADQCQKATEKAQE-----KRIHQP 132
DB 474 EDSPVDD-----MFGSPNEAATAKSDQEDQETPVTNTLDFP 512
QY 133 LPLAFE-----SQTHSFQSQKQOVLSQQLP-----FIFOSSQAKSESVMK 174
DB 513 LSVANDYTQPTAGPSDDMFGFCNSADLPQ--TDSVPHSGATADSMFGFDSVPATGASVEP 570
QY 175 HTSVIQSTGLSVNVKPLSVNQAKETVMKLIVPSPDVLKAGNKNYSESSLLTSELRSK 234
DB 571 PTETPELHD-ASQTOP--TLDMPITPEHIPTIEDNQDLSHESQNESQVASESOLRET 627
QY 235 REYKQAFPSPQKKQESSKSLKKVIAA-----LSNPKATSSSPAHPKQPL 279
DB 628 NE-----TLEGEEMKELAQEPAVEDMFGCGDASAPVEDSSSTEKATESHTPPNLA 679
QY 280 ENNHNPFLTNALLGNHQPNGVIOSVIOEAPLALTKTKMOSKINENIAAASSTFFSP- 338
DB 680 D--LPPPVASASLL-----PVQTLSEAKQSAARAATTPVADMAIDIDSMISAAALVPQPSA 733
QY 339 ----VNLSTSGRRPTGNOTPWPSPASPLHSGQKEKAVNNVNPVKTQHHSPAKSLVQ 394
DB 734 PEQKVEDQVLDESSPNVITABLP---PIVVKGSLEEA--NDKSAAESQHEP-----VE- 781
QY 395 FRGTDSDTPSKDSDSNED---EEDDEDEDEDEDEDESD--SQESDSNSSDYEG 449
DB 782 -LSEDRDL--SKDELVAQDPLPAEDTDQDFVEPAVPVEALERLAISETFKLEELQF 838
QY 450 SEEDDDDDKQDESDDSTEGEKTSMKNKNTSSVKSMSLTGHS-TPRNLIHIAKAGSA 508
DB 839 TYLMPEPNREELDKLQMTDHEQQLIQLKDEAERLQRENKVLAGRSAPRHLH----- 890
QY 509 PAALCSSESQAPFLGTSSTLTSPPHS-GTSKRRTVDERELRIPLYGQWQRETRINFG 567
DB 891 PNEFDSKISVSFF---EGLNNEPHTSDGIS---LTTGEELRL----- 927
QY 568 GRLOQEVAYAPCGKKLRQYPEVIKYLRSNGIMDISRDNFSPSAKIRVGDYFEARD---- 623
DB 928 -----LALEQGD-YEALDALHG 944
QY 624 -GPOEMQWCLLKEDVPIPRAMEGRRRGRPPNPDRQR-----AREESRMR 668
DB 945 MSPEELQMEQREERFNKLKELTDRRDELLYKDDTEALKLIEBIWAFVDAEESKMR- 1003
QY 669 RKGRPPNVGNAEFLDNADAKLRLKQAEIARQAQAIKLLRLKQEQEAVAKKQQA 728
DB 1004 -----IQAADLLKGAQEEDERRA---EIEAKI-KEEQDRAEERLQOE 1046
QY 729 IMAAEEKKEKQ---KEQIKIMKQEKIKRIQIIRMEKELRAQIILEAKKKKEE-AAANAK 783
DB 1047 MIALQKQOQYAEQETKADLLREAMEKQERIEEQRIK--AEARRLLIDDEATAE 1104
QY 784 LLEAKRIKEKEMRRQAVLLKHQERRRRRRQHMMMLKAMEARKK--ABEKERLQEKRDE 841
```

```
Db 1105 RAAAEADIRVKALEEMARIKKELEBAAR-----TASEKRKLLKQKEQLEMAET 1157
QY 842 KRLNKRKLEQRRLEWAKELKKNEDMCLADQKPLPRLPGLVLSGSTFSDCLM-- 899
Db 1158 ARHHEE-DVAREHKELOTLHEERRKVRQORLASSEPMDELLREDGNDASADPFSYASLDI 1216
QY 900 ---VVQFLNFKVLGFDVNDVNPILSVIOEGLLNGDSMGVQDOLLVRLLSAAVCDPGL 956
Db 1217 ERIIREMKAQLEKEMTERKMALEHADNLAQEL--AAKGLSTLQ--AERLJSTAASDT-- 1270
QY 957 ITGYKAKTALGSHLNVGNRDNVSEIL---QIFMEAHCGQPTFESLKTKAFQAHTPAQ 1013
Db 1271 ---VARLDNEDRLFNEAMDARQELLKKEQGLVHQNESQRAHLEEQ-----R 1318
QY 1014 KASVLAFLINELACSKSVSEIDKNIDYMS---NLRDKWVV---EGKRLKRLRIHAKKT 1067
Db 1319 KAKKLARLVEQVAELAEEKHNLRIKIAHGNKEFHKKESKLVASNGKALKARKEHAKEH 1378
QY 1068 GKRDTSGGIDLGEEOHPLGTPTPGKRKRK-----GGDSYDDD---DDDDSDDOG 1115
Db 1379 EALETQ-----LSSVSEQURRAKQKLVNPOGTSMDORSVVDYLRKSTDRS 1427
QY 1116 DEDEDE-DEEDKEDQK-----GKKTIDICEDEDEGD-----QAASV-- 1148
Db 1428 EKEQVSAADKDHSPRLSVPEGSGISQPSLVSGHIGITAMSTGYVDHGGGKQHASDDL 1487
QY 1149 -----ELEKQLEKLSKQSOQY---RRKLFDASHLSRSMFGP-----DRYRRRWYI 1192
Db 1488 PVALOKLEHARLEKTEQRAAFASEREQLQAQHALEAQLSPVNVPSKLTRE-- 1545
QY 1193 LPRCGGIFVEGMESEGLEIEAKEREKLEKKAESVQIKEMPE-TSGDSINCSNTHCQK 1251
Db 1546 -QRKQIIFAE-----NNREKUSKAHNVKQIPKTIIDDILNERRKKGHGER 1593
QY 1252 EDLKEKDNNTNPLQKPGFSKLSKLEVAKMPSEVMTPTKNAGANGCTLSYQNSGKHS 1311
Db 1594 EQLK-----NMAVRTSAVDR-----HSRVSNVMDSSDDEWNDLAARTSSRYA 1636
QY 1312 LGSVQSTQTSQVNEKADSNL---FNTGSSGP 1340
Db 1637 DISAKPRSQAOVQFASDRIDPAFNSYTEAP 1668
RESULT 48
Q8MKZ9 PRELIMINARY; PRT; 460 AA.
AC Q8MKZ9;
DT 01-OCT-2002 (TReMBLrel. 22, Created)
DT 01-OCT-2002 (TReMBLrel. 22, Last sequence update)
DT 01-WAR-2004 (TReMBLrel. 26, Last annotation update)
DE CG10897-PD.
GN Name=tou; ORFNames=CG10897;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxId=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brannon R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahle C., Davenport L.B., Davies P.,
```

```
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin C.J., Evangelista C.C., Ferraz C., Ferrieria S., Fleischmann W.,
RA Foster C., Gabriellian A.C., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
RA Jalali M., Kalueh F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E.C., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA Williams S.M., Woodgett, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RL "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=2426065; PubMed=12537568;
RA Celniker S.E., Wheeler D.A., Kronmiller B., Carlson J.W., Halpern A.,
RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,
RA George R.A., Hoskins R.A., Laverty T., Muzny D.M., Nelson C.R.,
RA Pacleb J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,
RA Svirskas R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
RA Weinstock G., Scher S.E., Myers E.W., Gibbs R.A., Rubin G.M.;
RT "Finishing a whole-genome shotgun: Release 3 of the Drosophila
RT melanogaster euchromatic genome sequence."
RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426070; PubMed=12537573;
RA Kaminker J.S., Bergman C.M., Kronmiller B., Carlson J., Svirskas R.,
RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,
RA Ashburner M., Celniker S.E.;
RT "The transposable elements of the Drosophila melanogaster euchromatin:
RT a genomics perspective."
RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426069; PubMed=12537572;
RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
RA Bettencourt B.R., Celniker S.E., de Grey A.D., Drysdale R.A.,
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA Lewis S.E.;
RT "Annotation of the Drosophila melanogaster euchromatic genome: a
RT systematic review."
RL Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).
RN [5]
RP SEQUENCE FROM N.A.
RG FlyBase;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE FROM N.A.
RG FlyBase;
RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE003825; AAM68714.1; -.
DR HSSP; Q14839; IMM3.
DR FlyBase; FBgn0033636; tou.
DR GO; GO:0003677; P:DNA binding; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
```


RP SEQUENCE FROM N.A.
 RC FlyBase;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AE003686; AAF54511.2; -;
 DR FlyBase; FBgn0037800; CG3996.
 DR InterPro; IPR001005; Myb DNA binding.
 DR InterPro; IPR000195; RabGAP_TBC.
 DR Pfam; PF00566; TBC; 1.
 DR SMART; SM00164; TBC; 1.
 DR PROSITE; PS00037; MYB 1; UNKNOWN 1.
 DR PROSITE; PS00086; TBC_RABGAP; 1.
 SQ SEQUENCE 3111 AA; 349635 MW; EDA9A5FD38115773 CRC64;

Query Match 3.6%; Score 365.5; DB 2; Length 3111;
 Best Local Similarity 18.8%; Pred. No. 1.8e-05;
 Matches 397; Conservative 314; Mismatches 734; Indels 669; Gaps 97;

QY 2 GOTKSTSGGNRK-----CHQEQSKNQPLDARVDKIKDKKPRKKA-MESSNSDSDSGT 55
 Db GEQESGTGGGSGSTDTSLCDDDDPKS-----TEKSPKQAKLARKLKEQQLAG 665

QY 56 SDDTSSEG-----ISSDSDDLEDEEEDQSIIE-----SEDDSDSSE 96
 Db SRETSLEQRPRKSWAPSSHEIPFMLMGDTSGD-----EKEDKSTKEGPIGTQDAEDSATE 720

QY 97 AQHKSNNQVLLHGIS-----DPKADQ-----KATEKAQEKI--HQPLPL 135
 Db SGHYEFDE-LHLVSSKMEPLKLPDPDEFTGMTSVSPITPREKSEAEEDLLDERKPPDV 779

QY 136 APESOTHF--OSQOKPOVLSQLPFIQSSQA-----KEESVNHKT----- 176
 Db SDDGVTNQYFERVNSVERPNRL--ELTYSLNEEETDTNAIYLEEREKVEGSHGDBREYNSL 837

QY 177 -----SVIQSTGLVSNVKNPLSLVNOAQKETYMKLIVSPDVLKAGNKT-----SE 222
 Db PPFPYRENDGQGGGKVPQIRDDNIPGE-NKDDYKELLSMTIE-----ENTVYKPTPT 890

QY 223 ESSLLTSLRSKREYQKAF--PSOLKQESKSLKVIKVAALSINPKATSSSPAPK----- 276
 Db TASTLSNASRKRDRPRRTLTARSSTIEIEERYQALERRISQ-DQPSGDRQAKYIPSTAAL 949

QY 277 ----QTLNHNPNPLTNALLGNHOPNGVIOQVIEAPL-----ALTTKMQSKNINIAA 329
 Db EERFNTLEKQ-----LSAEQRKELEMEAEYPIKSERIPSTADLESRFNSLTQ 999

QY 330 ASSTPFSS--PVLNLTSGRRTPGNOTPYMPASPIHLSQGEKAVSNVNVNPKVTOHSHHP 387
 Db MSSSESSSKTPIDLKDEK-----PSGSSSKNQKDESKTSKLH---KSEEPESN 1045

QY 388 AKSLVEQFRGTDSD-----TPSSKDSQED-----SNEDE 415
 Db TKETTGETEASDSDSKIGEKETEOPRIKLPSTAELEDRFNALERKMSVQKSPSKNKK 1105

QY 416 EEDDEEEDDEDDSDSOSDSNSESDETEGSEEDDDDDKQODESDSDTEGEKTSMK 475
 Db EPPDEESKSTKEPEPESEKANEKTSGRQTPIAKK---DSKSDQKSETKENQSTPK 1162

QY 476 LNKTTSSVKSPSMSLTGHTSTPRNLHIAPGAPAPALCSQSAPFLGTSSSTLTSSPHS 535
 Db NQDEKVKVKSP-----KSEEMIEKETSSNPKEDSHES 1194

QY 536 GTSKRRRTVDRBELRIPLEYHQRETRIRNPGGRLOGEVAYAPCGKKLRQYPEV----I 591
 Db EAATNKKEVGNREL--SSEKG---DHKTK-----EKSEAP--GKAGKETAETKNANV 1240

QY 592 KYLSRNGIMDTSRDNFSPSAKIRVGDFFEARDGPQEQMOWCLLKEDDVIPIRAMEGRGR 651
 Db KDSSKKG--DSQKNE--AAKTSVSQ-----TESDL-----K 1267

QY 652 PPNPRQARERESRRRGRPPNVGNAAFILDADAKLLRLQAOETIARQAAQIKLRL 711
 Db PSSKENSTSKDAEQEKTPRKSPSPSTELEKRFNALQKQMSWTTL-ETTKPEQDTKPKTS 1326

QY 712 QKQEQARVAKKQQAQIAAAEKKRKQKQEIIMKQOEKIKRIQOIRMEKELRAQQILEA 771
 Db TSAAEVTKQSKMKSPPDDIKIE-VNVAIEKQS-----RVEVEVNA-----EK 1368

QY 772 KKKKKEEAAAKLJEAEKRIKEKEMRRQQAALLKHQBERRRROHMLMKAMEARKKABEK 831
 Db KRKNVEEAP-----KNKEGDSQQPESQHKGNQRR-----ASEPSTEDLE 1410

QY 832 ERLKQEKREKRLNKKERLEQ-----RRLENAKELKKPNEDMCLADQPLPPLPRIP 885
 Db KRYETLKR--RMSSKNQFSETVDEALERIQOEVRSE-----AVEBEKPPSTEDLE 1459

QY 886 GLVLSGSTFSDCLMVVQFLRNFGLFDVNI DPNLSVLQEGLLNIGDSMGEVQDILLVR 945
 Db S-----RFEALHGDKKNVSKSMDETKHVDVA 1485

QY 946 LLSAAVCDPGLITGYKAKTALGEHLNVGNRNVSEILQIFMEAHCCQOTELTSLTKTKA 1005
 Db IEAHIPSPPPPPPPPPKRPVLAEPVLH---QQQALIEELQSKMR---GQSPGEENLKPS 1539

QY 1006 FOAHTPAQKASVLAFLINELACSKSVSEIDKNIDYMSNLRRDKWV----- 1051
 Db INPQRRQKK-----LLQRPPTMGDETSEAPANTAYTRAANHEQWQQRVMVRRFSDLP 1593

QY 1052 -VEGKLRKLRIIHAHKTGKRTSGGIDLGE--EQHPLGTPTPGRRKREKGGSDYDDDD 1108
 Db DLENRLQFLERQLYKFKYQKRCASDEVASRVKLPDPEDQSTSRQARKQ----- 1642

QY 1109 DSDSDQDDEDEEDKEDQKGTDTICEDDEDEGQAASVEELEKQTEKLKSKQSQYRRK 1168
 Db -----EAEG-----QLQORVLALEKQLSNSLK 1665

QY 1169 LFDASHSL-RSMFGPDRYRRYWLPRCGIIFVEGMEGEGLEIEAKERE-----KL 1220
 Db LLEAMRHRHSADSGSPRLSTETIDATGKELVRYTQNI GELSEVDAHKFINISINIKM 1725

QY 1221 ---KKAESVQIKEMFFETSGDSLNCSTNTHCEQEKD--LKEKDNTNLFLOKPGSPSKLSK 1275
 Db MNKDSKSKPKGSKPTEDL-----TRRLQLEQOLLEER-----AKNGS----- 1767

QY 1276 LLEVAKPPSEVMVTPKPN--AGANGCTLSYQN--SGRHSLG-----SV 1315
 Db -----IPPEVLEEKPEKLEEKDCKQKKNCHQHVKGDEVEKTEIPADRKIEPASA 1821

QY 1316 QSTATQSNVEKA-----DSNNLFNTSGSGPGKFFYPLNDQLLTKLTENR 1361
 Db KETKTLENVEKAQTRAKVVDTEKSVKQDNQAVTDEKS-----VQDQVNVVDDKKADR 1871

QY 1362 QWFSLLPRTPC---DTSLTTHADMSTASLVTPOSQP-----SKSPSPTPAPLGSSAQN 1413
 Db KILDKKSPAGKSEDTKQTSKKKESDIKQASEAPKAGASKETSTRGPSETKLEKP 1931

QY 1414 VGLNPFALLPQKQGVGVMGLQFCGMPGTGVYVTSNIPTFLSVPSLGSGLSGSEGNNSFL 1473
 Db T-----TKESVLKETFP---KKENLESEKPKSEKENE---A 1960

QY 1474 TSNVASSKSESPV-----PONEKATSAQAAVEVAKVPDPPSPKPIPEMQGFWRIIDPE 1529
 Db TKTQTSKETPTTAVASPKESKVSQK--MTEKKEKTIKDSSSKELPEKM-----VINST 2012

QY 1530 DLKALLVHLRGIREKALQKQIQKHLDIITQACLKNKDVAIIELNENEENQVTRDIVEN 1589
 Db DVGMDP-----NGKTVVLLMDNEHRASKVR--LUTR 2042

QY 1590 WSVEE-----QAMEMDLSVLOQVEDLERRRVASLQVGMWCPPEPASEREDLVYFEHKS 1644
 Db ANTEELEDLFQALEKQLNDRNLVKSDEGLRIVD-----PKPSAQVE---QTOAI 2090

QY 1645 TKLCKEHDGETG--EDESSAHALERKSNPLD-----TAVTRLA 1682
 Db SDLTKEIE-DPTSAKPEENPKAAEKDPEPEPFDWGPNTVKKHLKRTKTVLPSTK 2149

QY 1683 DLE---RNRIERIE--EDIAPGLRVWRRLSEARSAAQVALCIIQOLQKSI AWEKSIKMYV 1737

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 9, 2005, 13:38:21 ; Search time 36.9823 Seconds
(without alignments)
4885.990 Million cell updates/sec

Title: US-10-702-148-13
Perfect score: 9826
Sequence: 1 MEMEANEANDHFNFTGLPPA.....MRRFFESRWEEFYQKQANL 1878

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 65 summaries

Database : PIR_79:.*
1: pir1:.*
2: pir2:.*
3: pir3:.*
4: pir4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1021	10.4	449	T12495	hypothetical prote
2	863.5	8.8	1430	T34516	hypothetical prote
3	538	5.5	1479	T17401	transcription regu
4	441	4.5	811	T08738	hypothetical prote
5	344.5	3.5	1051	S55259	TIF1 protein - mou
6	292.5	3.0	2187	T30826	nascent polypeptid
7	286	2.9	2464	1 QRMSP1	microtubule-associ
8	286	2.9	2526	T20531	hypothetical prote
9	285	2.9	2722	T20532	hypothetical prote
10	284	2.9	5327	T13564	microtubule-associ
11	280.5	2.9	2374	A36577	microtubule-associ
12	280.5	2.9	2774	A43359	microtubule-associ
13	279	2.8	2738	T20531	protein F07A11.6 [
14	272	2.8	3968	T44265	trithorax homolog
15	258.5	2.6	4957	T03455	ALR protein - huma
16	258.5	2.6	5262	T03454	ALR protein - huma
17	253	2.6	2359	T03094	A-kinase anchor pr
18	250.5	2.5	1502	T34513	hypothetical prote
19	248	2.5	1824	1 QRHMT	microtubule-associ
20	247.5	2.5	3924	2 S37431	ankyrin 2, neurona
21	245.5	2.5	2649	2 T51023	hypothetical prote
22	244.5	2.5	6642	2 T29757	protein UNC-89 - C
23	243	2.5	1819	2 T32008	hypothetical prote
24	242	2.5	3942	2 T42730	Bassoon protein -
25	241.5	2.5	1234	2 T00363	hypothetical prote
26	240.5	2.4	3869	2 A48205	All-1 protein +GTE
27	240.5	2.4	3938	2 T42761	Bassoon protein -
28	240	2.4	1299	2 T47182	hypothetical prote
29	238.5	2.4	1400	2 T52359	hypothetical prote

30	238.5	2.4	1830	2	A37981	microtubule-associ
31	238	2.4	1825	2	S13507	microtubule-associ
32	237.5	2.4	3256	2	A48666	cell proliferation
33	233.5	2.4	1279	2	T13613	hypothetical prote
34	230.5	2.3	1551	2	F86342	F9H16.4 protein -
35	230	2.3	3488	2	T34418	hypothetical prote
36	229	2.3	5762	2	A41819	proline-rich pepi
37	228	2.3	2282	2	T42717	DNA-binding protei
38	227.5	2.3	2845	2	T49505	adenomatous polyo
39	227	2.3	7962	2	T38346	elastic titin - hu
40	224.5	2.3	1320	2	JC5630	TCOF1 protein - mo
41	223.5	2.3	1046	2	T42720	cytoplasmic linker
42	223	2.3	606	2	A43427	neurofilament trip
43	223	2.3	1378	2	G88637	protein F53H1.4 [i
44	223	2.3	2843	1	R8HUAP	adenomatous polyo
45	220	2.2	2897	2	B48666	cell proliferation
46	219.5	2.2	1634	2	T26517	hypothetical prote
47	219	2.2	1046	2	T42734	cytoplasmic linker
48	217.5	2.2	918	2	C96829	unknown protein F1
49	217.5	2.2	2140	2	T18543	probable cell-adhe
50	217	2.2	877	2	I50591	class II INCENP pr
51	216.5	2.2	1805	2	A34736	neatin - rat
52	216	2.2	2453	2	S60254	nuclear receptor c
53	216	2.2	2688	2	T49477	alpha-A-crystallin
54	216	2.2	3381	2	T42389	versican precursor
55	214.5	2.2	5170	2	T15348	hypothetical prote
56	214	2.2	1606	2	T49219	translation initia
57	212.5	2.2	1596	2	A35927	190K DNA-binding p
58	211	2.1	1396	2	A44453	translation initia
59	211	2.1	1828	2	A40115	microtubule-associ
60	210.5	2.1	854	2	S02003	neurofilament trip
61	210	2.1	1560	2	T42727	proliferation pote
62	209.5	2.1	1020	1	QFHUH	neurofilament trip
63	209.5	2.1	1804	2	T34518	neatin - golden ha
64	209	2.1	839	2	I50590	class I INCENP pro
65	208	2.1	681	2	T17342	hypothetical prote

ALIGNMENTS

RESULT 1

T12495
hypothetical protein DKFZp434H071.1 - human (fragment)
C:Species: Homo sapiens (man)
C>Date: 23-Jul-1999 #sequence_revision 23-Jul-1999 #text_change 09-Jul-2004
C:Accession: T12495
R:Poustka, A.; Klein, M.; Mewes, H.W.; Gassehuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, June 1999
A:Reference number: Z17525
A:Accession: T12495
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-449 <POU>
A:Cross-references: UNIPROT:Q9UIF8; EMBL:AL080173
A:Experimental source: adult testis; clone DKFZp434H071
C:Genetics:
A>Note: DKFZp434H071.1
C:Superfamily: transcription factor GCN5; bromodomain homology
F:366-421/Domain: bromodomain homology <BRO>

Query Match 10.4%; Score 1021; DB 2; Length 449;
Best Local Similarity 41.4%; Pred. No. 6.6e-39;
Matches 206; Conservative 74; Mismatches 127; Indels 90; Gaps 8;

Qy	1415	IPDEMLDAMLKALHGRIGREKALHKLHKLHGRIFLOEVLRLPSADPIPEPQLPAFQ---	1471
Db	2	IIDPEDLKALKLVHLRGIREKALQKIQKHLDTQACLKNKDVAILTELNEENQVTR	61
Qy	1472	EGINWSWPKEITYETDLAVLQWVEELRQVIMSDIQRTGWTCPSPDSTREDLAYCEHLSD	1531
Db	62	DIVENWSVEEQAMEMDLVLOQVDELERRVASASLQVKGWMCPEPASREDLVVFEHKSF	121

QY	1550	KTTNPDLDAWRLA-----ALEQNKKRYRIEPLWPHTHEVVLEKALLSTNGAPEGT	1502
Db	995	KPIVPPTMALAQIVKDMDMAWKVIDEEVGQELDE-----TIIRQKII-----E	1038
QY	1603	TTEISYEITPRIRWRTOQL-RCRSAAHVCLCLGHLERSIAWEKSYNKVTCLVCRKGND	1661
Db	1039	TAMDQDTCOLFDWKSYVSTEATQTSQLMVALQTLEGIMWRSREALCQCKSMGDG	1098
QY	1662	EFLLLCDGDRGCHYCHRPMKEAVPEDGMFCTVC-----	1696
Db	1099	E-MLVCDGESGCHMECFRPMTKVPEGDFWCQRCEEKSGRPMCMPCSRRETGNLHCOR	1157
QY	1697	LAOQVEGEFTQ-----KGFPPKR-----G	1715
Db	1158	CAYHVHOCSDQGPKAEINPETTICGHCBQMKNRFVKRLILRSESEERELEDNHAENG	1217
QY	1716	QKRGSGYSLNFSEGDRRRRVLLKGRSPAAGPRYSERLSPSKRRRLSM-----RNH	1768
Db	1218	ENTKNHG-MMGNG-----ATAIGVINQQGVGNLKRLKLEVPSIGGLPKNM	1263
QY	1769	HSDLTFCEIILMEMESHDAAPFXEPVNPLVSGYRRIIKNPMDFSTMRELRGGYTSS	1828
Db	1264	NKEL--COLMLDELIVQANALPLEPNVPLVGYKMIISKPMDLKTIRQKNEKIYETP	1321
QY	1829	EEPAADALLVFNCOPTNEDDSEVGKAGHMRRFFESRWEE	1869
Db	1322	EDFAEDIELMFANCQFNIDHSBITGRAGISLHKFFQKRWFQ	1362

RESULT 3
 Tl7401
 transcription regulator WBSCR9 - mouse
 N;Alternate names: Williams-Beuren syndrome deletion transcript 9 homolog
 C;Species: Mus musculus (house mouse)
 C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
 C;Accession: Tl7401
 R;Peoples, R.J.; Cisco, M.J.; Kaplan, P.; Francke, U.
 CytoGenet. Cell Genet. 82, 238-246, 1998
 A;Title: Identification of the WBSCR9 gene, encoding a novel transcriptional regulator,
 A;Reference number: Z18735; MUID:9907764; PMID:9858827
 A;Accession: Tl7401
 A;Status: preliminary; translated from GB/EMBL/DDDBJ
 A;Molecule type: mRNA
 A;Residues: 1-1479 <PEO>
 A;Cross-references: UNIPROT:Q9Z277; EMBL:AF084480; NID:G4165088; PID:G4165089; PIDN:AADD
 F;1360-1415/Domain: bromodomain homology <BRO>

Query Match 5.5%; Score 538; DB 2; Length 1479;			
Best Local Similarity 18.8%; Pred.No. 1.6e-16;			
Matches 352; Conservative 249; Mismatches 624; Indels 646; Gaps 70;			
QY	141	HNTNLRAQSOKFWANGTHSPMGLNF-----DSQE----LYDSPFDQNFEEVCSGIHPDEA	191
Db	38	YEARLERYSERIWTCKSTGSQLTTHKEAWEEGEVAELLKEEFNYEKLVLVEMVHNTA	97
QY	192	AEEKMT-----SVVAENGTGLVCSLEL-EXQPELKMCGVNGSVPSVESLHQEVSLVDP	245
Db	98	SLEKLVDASWLMEINTKYAVGEEDCFYEGKMKLVKITV----KHIFLEKVDDEA---VEK	150
QY	246	PTVSCLDPRSHLPDQLEDTPILSDESLPEFPNSLAPEVPVSGLYCIDD-----TEL	295
Db	151	KSGACDSPS---SDKENSQAQD-LQKETVVKED-EGRRESINDARRSRPKUPTS	205
QY	296	MGAEDK-----LPLXDSPVISALDCPSLNNAFAFSLLADDSTTSIFSAPSPVVLGE	349
Db	206	KGERKWAPPKFLEP-----HKYDVKLQNEDKLIISNPADSLIRTE---RPNKAIL-R	254
QY	350	SVLQDNISFDLINGSDAE---QEEMETOSSOFFPSLTPQAPDQSQSTTLQHLPATSPAVPTT	406

Db	255	YFIRHNALRAGTCENAPWVVEDELVKYS-----LPSKESDFELDPYKWTMLNPST	305
Qy	407	SPAVSLVVSPPAASPEISPEYCPAASTVSPAVFSVSPASSAYLPAVLSLEVPITASVTSP	466
Db	306	KR-----RNTGSPDRKPSKKPRDS-----SSLSSPLNPKLWCHVHLEKSLNGP	349
Qy	467	-----KASPVTSPPAAAFPTASPANKDVSSFLETTADVEITCEGLTASGSDVMRR	517
Db	350	PLKVYNSKNSKSPBEHLEGVMKIMSPNNKNLHSP-----HIPKGPAAK-----	393
Qy	518	RIATPEVRPLQHGWRREVRIKKGSHRWGETWYYPGCGKRMKQKPEVTKYLSRNLVHS	577
Db	394	-----XPGKH-----SDPLKAKGKGKILNG	415
Qy	578	VREHFSFSRPMVPDGFEEEDTPEGLWVQLSAEBIPSRQIAIT-----GKRGRPRNTEK	633
Db	416	QKSTGNSKSPSKCV-----KTPK-TKMQMTLLDMAXGTQKWTRTPRSSGGVPRSSGK	467
Qy	634	AKTKEVPKVKRGGRPPKVIITELLNKTNRPLKK-----LEAQTLEEDKAKIAKS	686
Db	468	PH-KHLP-----PALHLIIAYYKENDKEDKSALSCTVSKTARLLASNEBARLPPEE	518
Qy	687	KKMWRQKVRQEGCLTTIQGARNKRQETKSLKHKE--AKKXSAXEKKGKTQKQELKEK	744
Db	519	LRALVQK-----RYELLSHKRKNWASMBEQKRYVKKRQBEKER	558
Qy	745	VKREKKEKVKWKEEYTKAKPACKADKTLATQRRLEERQKQOQMIEMKKPTDECLTD	804
Db	559	LREKAKER--REREMLER-----LEKQRFEDQE-----LG	587
Qy	805	HQBLPDPFSRVPGLT-LPSGAFSDCLTIVFLHSFGKVLGFDPAKDVPSLGVLOEGLLCOG	863
Db	588	GRNLPAPRLVDTPEGLENTLFGDVALWVEFLSCYSGLLLPDAQYPTAVS-LMEALSADK	646
Qy	864	DSLGEQDILLVRLKAAHDPGPPSPQCSKLILGEKVSEIPLTRDNVSEITLRCFLMAYG-	922
Db	647	GGFLYNRLVILQTLQDEIAEDYGE---JGMKLSIPLTLHSVSELVRLCLRRCDV	702
Qy	923	-----VXPALCDRLRTQFPQAPPOQQAVALVAPFVHELNGSTL	960
Db	703	QEDSEGETDDNKDSTPFEDNEVQDEFLEKLETSFEFELTSEBKAILTALCHRI---L	758
Qy	961	IINEIKTLESMSYRNKNKWIVEGRRLRK-----TVLAKRTGRSEVEM-----GRPECL	1011
Db	759	MTYSVDHMETROQVSAELW--KERLAVLKEENDKKRAEKQKKEARNKENGKEENVL	816
Qy	1012	GRRRSRIMEETSGMBEEEEESIAAVPGR-----GRDGEVDATASSIPELERQIELKL	1066
Db	817	GKVDKKEIKVIEQVEVEADDMISAVKGRRLLSMQAKKREIQERETKV-RLEREABEE	875
Qy	1067	SKR-----QLFPRKLLHSSOMLRAVSLQDQRRYRYWLP-YLAGIFVFGTEGNLVPE	1119
Db	876	RMRKHAAAEKAFQEGIAKAKLVRLTPITGTDENHNRYWLFSNEVPGLFIE-----	926
Qy	1120	EVIKKETDSLKVAHAASLNPALFSMKWELAGSNVTTASSPARAKSRPLKTKPGFMQPRHFK	1179
Db	927	-----KGWVHNSID-----YRFKXHR-----	935
Qy	1180	SPVRGQDSEQPAQLQPEAQLHVPAPQPOLQLQLOSHKGFLEQEGSPSLSGOSHDLSQ	1239
Db	936	-----YRFKXHR-----	942
Qy	1240	SAFLSWLSQTSQSHSSLLSSSVLTPDSSPGKLDPAQSPQPEEPEDPEABESSPOLQAFWFI	1299
Db	943	-----KQHSNL-----PDD-----DYCPRKKA-----NL	962
Qy	1300	SAQWPCNAAPTPPLAVSEDOPTSPPOOLASSKPMNRPSPAAANPCSPVQFSTPLAGLAPKR	1359
Db	963	GKNASVNAHHGPALAEV---TTVPKQ-----	986
Qy	1360	RAGDPQEMPOSPTGLQCPKRGRRPPSKFFKQMEQRYLTQLTQAQVPPMCSGHWIWPDPPE	1419
Db	987	-----GQ-----NLWFLCDSQ-----	997

```
Qy 1420 M-LDAMLKALHPRGIREKALHKLHNRDIFLQEV--CLRPSADPIFPRLQPAFQEGIMS 1476
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 998 KELDELLSCLHPQIRSQPKERLEK---YQETHCNMARKNVLGKSCDGNQOE-LLN 1053
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 1477 WSPKE-----KTYETDLAVLQWVEELEQRVIMSDLQIRGWTCPSPDSTREDLAYCEHLS 1530
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1054 FLRSDLIEVATRLQKGLGYMEGTSEFEARVIS-----LEKJK 1091
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 1531 DSQEDITWRGCGREGLAPORTKTTPLDLAVNRLAALAEONVRRYLRPLWPTHEVLEKA 1590
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1092 DFGCV-----IALQASVTKFLQGFMAFKQ---KRK 1121
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 1591 LLSTPENGAGEPTTTEISY---BITPRIRIWRQLQRCRSAAHVCLCLGHLERSIAWEKS 1646
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1122 LQSEDSTKSEVDEKQWBEAKVASALEKWKTAIREAQTSRHHVLLGLMDACIKWMS 1181
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 1647 VNKVTCVCRKGDNDEFLLCGDRGCHYYCHRPKMEAVPEGDWFCVCLAAQVGEFT 1706
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1182 AENARCKVCRKGGDDKLILCDECNKAFHLFCLRPALYVEVDGEMQCPAC----- 1231
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 1707 QKPFPK--RQCKRKGYSLNFSGDGR-----RRRVLLKG 1740
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1232 QPPTARNRSRGNTYESTSGSEGDSEGESEEEEEEEEDYEVAGLRLRPKRTIRG 1291
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 1741 RES--PAAGP--RYSEERLSPSKR-----RRLSMRHHSDLTFCETILME 1781
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1292 KQSVIPAARPGRPCKKSHARRSRPKDDPEVDLVLQTKKIS--RRSLELOKCEDILHK 1350
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 1782 MESHDAWPFEPVNPRLVSGYRRIIKNPMDFSTMRELLRGYTSSEFAADALLVFDN 1841
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1351 LVKRFSPFPREPVTROEAEDYDVIEHPMDFQTIQNKCSGNYRSVQEFLLDMKQVFAN 1410
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 1842 CQTNEDDSEV 1852
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1411 AELYNCRGSHV 1421
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 4
T08738
hypoetical protein DKFZp586E0518.1 - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 09-Jul-2004
C;Accession: T08738
R;Wambutt, R.; Heubner, D.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, March 1999
A;Reference number: Z16471
A;Accession: T08738
A;Molecule type: mRNA
A;Residues: 1-811 <WAM>
A;Cross-references: UNIPROT:Q9NRL2; EMBL:AL050089
A;Experimental source: adult uterus; clone DKFZp586E0518
C;Genetics:
A;Note: DKFZp586E0518.1
F;709-764/Domain: bromodomain homology <BRO>

Query Match 4.5%; Score 441; DB 2; Length 811;
Best Local Similarity 20.8%; Pred. No. 1.7e-12;
Matches 166; Conservative 103; Mismatches 261; Indels 270; Gaps 18;

Qy 1302 QMPCNAAPTPPLAYSDQPTSPQOLASSKPMNRPANPCPS----- 1343
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 14 QEQINCVTREPLTDADEBALQKHQKKELEKIQSAIACTNIFPPLGRDMRYRYWIFP 73
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 1344 --PVQFSSTPLAGLAPKRRAGDPGEM-----PQSPTGLGQPKRRGRPPSPKFKQMEQ 1393
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 74 SIPGLFIEDYSGLTEDMLLRPSPSQNNVQSPQVSTTKGPELW-----SESTSNIDQ 128
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 1394 RYLQTLTAQVPPEMCSGWVWIPDEMLDAMLKALHPRGIREKALHKLHNRDIFLQEV 1453
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 129 GPRDHSVQLPKVPKPNRWCFYSCEQLDQDQIEALNSRGHRESALKETLLOEKSRICAQL 188
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 1454 LRPSADPIF-----EPRQLPAFQEGIMS--NSPKTEY--TDLAVLQWVEELQRVIM- 1503
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

```
Db 189 ARFSEEEKFHFSDKPPQPSKTYSGRSSNAYDPSQCAEKQLERLRLDFLLDIEDRIYQG 248
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 1504 -----SDLIQIGWTCPSD-----STREDLAYCEHLSDSQEDITWRGP 1541
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 249 TLGAIKVTDRIHWRSALESGRYELLSEENKENGIIKTVNEDVBERMEIDQTKVIVKDRLL 308
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 1542 GREGLAPORTKTPN-----LDLAVMRLAALAEONVRRYLRPL-----WPTHEVVL 1587
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 309 GIKTETPSTVSTWASTPQSVSSVVHYLAMALFQIEQGIERRFLKAPLDASDSGRSYKTVL 368
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 1588 EKALLSTPNGAGEPTTTEISYETPRIRIWRQLQRCRSAAHVCLCLGHLERSIAWEKS 1647
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 369 DR-----WRESLLSSASLSQVFLHLSTLDRSVIWSKSI 401
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 1648 NKVTCLVCRKGDNDEFLLCGDRGCHYYCHRPKMEAVPEGDWFCVCLAAQ----- 1700
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 402 LNARCKI CRKKGDAENMVLCDCDRGHHTYCVRPKLTVPPEGDFCPCPRKQSRRLSS 461
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 1701 -----VEGEFTQKPGPKRGKRGYSLNFSGD-----GR- 1732
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 462 RQPSLESDEVDSEMGDEDDVDGDEEGQSSEEEYEVEQDEDDSQEEBEVSLPKGRP 521
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 1733 --RRRVLLKGR-----ESPAAGPRYSE----- 1752
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 522 QVRLPVKTRCKLSSPSSRQOQEPGYPSPRSQOSTPKTTVSSKTRSLRKINSAPPTET 581
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 1753 -----ERLSPSKERR----- 1762
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 582 KSLRIASRSTRHGHGPIQADVFFVELLSPRKRGRKRSANKTPENSPNPNFRVIATKSS 641
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 1763 -----LSMRNHS-----DLTFCEILMEME 1783
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 642 QSRSVNTASKLSQESKCRKRQSPSPVTLGRRSSRGQGVHLSAFELVVELV 701
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 1784 SHDAAWPFEPVNPRLVSGYRRIIKNPMDFSTMRELLRGYTSSEFAADALLVFDNCQ 1843
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 702 RHDDSWFFLKVSKIQVPDYDIKKPIALNIIRKVNKCEYKLASEFIDDIELMFNSCF 761
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 1844 TFNEDDSEVGKAGHIMRRFP 1863
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 762 EYNPNTSEAKAGTRLQAFF 781
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 5
S55259
TIF1 protein - mouse
C;Species: Mus musculus (house mouse)
C;Date: 27-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 16-Aug-2004
C;Accession: S55259
R;le Douarin, B.; Zechel, C.; Garnier, J.M.; Lutz, Y.; Torà, L.; Pierrat, B.; Heery, D.;
EMBO J. 14, 2020-2033, 1995
A;Title: The N-terminal part of TIF1, a putative mediator of the ligand-dependent activa
A;Reference number: S55259; MUID:95262642; PMID:7744009
A;Accession: S55259
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1051 <LED>
A;Cross-references: UNIPROT:Q64127; GB:S78221; NID:g999814; PIDN:AB34290.1; PID:g999815
C;Superfamily: bromodomain homology
F;48-136/Domain: RING finger homology <RRN>
F;925-981/Domain: bromodomain homology <BRO1>

Query Match 3.5%; Score 344.5; DB 2; Length 1051;
Best Local Similarity 21.0%; Pred. No. 5e-08;
Matches 209; Conservative 111; Mismatches 335; Indels 341; Gaps 45;

Qy 974 SYRKNVIVSGRLRL--KTVLAKRTGRSEVMGRPECLGRSSRIMETSGMBESEE 1031
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 262 AFONQKVIIDTLTKMEKTKYIKY-----GNQIONRIELNQKQVEQ 307
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 1032 EESIAAVP-----GRRGRDGEVDATASSIPELQRIEKLKSKQLFFPKLLHSSQMLR- 1085
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

Db 308 DIKVAIFTLWVEINKKG-----ALLHQLSLAKDH---RMKLMQOQOEYAG 351
Qy 1086 -----AVSLGQDR---YRRYRWLPY-LAGIFVGEFEGNLVPEEVIKKETD 1127
Db 352 LSKQLEHVMHFSAVSSGSSTALLYSKR--LITYRLRHLRLARCDASPTNTTIQFCD 409
Qy 1128 SILKVAHAASLALPASMKMLAGS---NTTASSPARASRLPKTKPGEWQPHFKSPVRG 1184
Db 410 -----PSFWAQIINIGLSLVIDKESQPMKONPVV----- 441
Qy 1185 ODSEQOQAQLOPEAQLHVPAPQ-----POLQLQSHKGFLEQEGSPLSLGQ----- 1232
Db 442 EOSSQPPCGGLPSNQLSKPTQLSLAQLRLQHQVMAQRQVQRRPAPVGLPNRMQGP 501
Qy 1233 -SQHDLSSAFSLWSLQTSQSHSLSSVLTPDSSPGKLDPAPOPPPEEPDEAESPD 1291
Db 502 IQQPSISHQHPRLINFQHSKNGPVLPP--YPOQLRYSPSQ--NVPRQTITKPNPL 557
Qy 1292 LOAFWNTLS-AQMPCNAAPTPLAYSEDOPTPSPOOLAS-----SKPM----- 1333
Db 558 QMAFLAQQAQIQQWQISSVOAPTATASSSSSTPSSPTITSAAGYDGKAFSSPMIDLSAPVG 617
Qy 1334 --NRPSAAN-PCSPQVPSSTPLAGLAKRRAGDPGEMQSPSTGL--GOPKRRGRPPSKF 1387
Db 618 GSYNLPSPDIDCS-----STIMLDNIARK-----DTGVDHAQP-----RPPSNR 657
Qy 1388 FQMEQRYLTQLTAQVPPMCMGWWIIPDPEMLDAM-LKALHPRGIREKALHKLHNR 1446
Db 658 TVQSPNSS-----VPSPLAGPVTMTSVHP----- 682
Qy 1447 DFLQEVCLRPSADPIFERPQLPAPQEGIMSWSPK-----EKYETDLAVLOWV----- 1494
Db 683 -----PIRSPASSVSGRSGSSSKPAGADSTHKVPVVMLEPIRIKQENSG 729
Qy 1495 -EELEQRVIM---SDLOIRGWTCPSPDSTREDLAYCEHLSDSOEDITWRGPGREGAP 1548
Db 730 PPENTDFPVVIVKQESDESRPONTNYPRSILTSLLNSSOSSAEEETVLRSQ-----AP 784
Qy 1549 QRKTTNPLDLAVMLAALQNVKRRYLREPLMPTHEVVLEKALSTPNQAGEGTTTETSY 1608
Db 785 DSTGQDP-----GLHQE-----NSSNGKSE----- 804
Qy 1609 EITPRIRIWRQLOCRSAAHVCLGHLERSIAWEKSNVKTCLVCRKGNDEFLLLCD 1668
Db 805 -----WSDASQ--KSPVHV-----GETRK-----EDDPNEDMCAVCQNGGE---LLCCE 843
Qy 1669 GCDRCCHLYCHRPKMEAVPEGDFWCTVCLAQOVEGEFTQKQF-----PKRGOKRSG 1721
Db 844 KCPKVFLHTCHVPTLTNPFSGEWICTFC-----RDLKPEVDYDCDVPSSHSEKRX-- 894
Qy 1722 YSLNFESEGDGRRRRVLLKGRSPAAGPRYSBERLSPSKRRRLSMNHHSLDTFCBIIIME 1781
Db 895 -----SEG-----LTKLTPIDRK-----CERLLLF 915
Qy 1782 MESHDAAWPFKEPVNPRIVSGYRRIIKNPMDFSTWRERLLRG--GYTSSEFAADALV 1839
Db 916 LYCHEMSLAFQDPV-PLTVDPDYKIIKNPMDLSTIKRLQEDYCMYTKPDEFVADFLIF 974
Qy 1840 DNCQTFNEDDSEVGKAGHIMRFFESRWEFFQHQK 1875
Db 975 QNCAEFNPDSEVANAGIKLESYFBEILLKNLYPEKR 1010

RESULT 6
T30826
nascent polypeptide-associated complex alpha chain, muscle splice form gp220 - mouse
N:Alternate names: alpha-NAC protein
C:Species: Mus musculus (house mouse)
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 09-Jul-2004
C:Accession: T30826
R:Yotcov, W.V.; St-Arnaud, R.
Genes Dev. 10, 1763-1772, 1996
A>Title: Differential splicing-in of a proline-rich exon converts alphaNAC into a muscle

A:Reference number: Z20889; MUID:96312450; PMID:8698236
A:Accession: T30826
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2187 <YOT>
A:Cross-references: UNIPROT:P70670; EMBL:U48363; NID:gi666688; PID:gi666689; PIDN:AAB187
C:Genetics:
A:Gene: Naca
A:Map position: 10
A:Introns: 24/1; 1996/1; 2024/3; 2050/3; 2093/3; 2142/3; 2183/3
A>Note: differential splicing converts alphaNAC into a tissue-specific DNA-binding activ
C:Keywords: alternative splicing; DNA binding; transcription factor

Query Match 3.0%; Score 292.5; DB 2; Length 2187;
Best Local Similarity 20.4%; Pred. No. 2.8e-05;
Matches 347; Conservative 173; Mismatches 674; Indels 503; Gaps 69;

Qy 16 GLP--PAPAAAGLKPSSGGLYNGSPMNFPOQKSLNGDVNVNGLSTVSHTTTSGIL 73
Db 466 GVPVSPLPATBGLKNLPIS--ALVNVGAPVSPAQAGLPTRKDTTLQPLAPIA-----L 516
Qy 74 NSAPHSSSTSHLHPSVAYDCLMNYSQVPSANPGSNLKDPPLLSQFSGGQYPL--NGILG 131
Db 517 KESPSQSASSL-----EVLSEDVTYKTKTGPPAPVVRPAIAG 554
Qy 132 GSRQPSPSHNTNLRASQK--FWANGTHSPMGLNFDQSQELYDFFPDQNF----- 179
Db 555 -----VATTTSLRADSPPAVIRADSCVSP--NTVSQPLKRSVTDPMAPRTAKNTAP 604
Qy 180 -----EEVC-----SGIHPDEAAEKMTSVVAENGTLVCSLELEX---QBEL 220
Db 605 STTSPLVLASEGCFVASSMALSPOASVSE-----TALALSPKIPKSPFPDPPL 655
Qy 221 KMCYNG-----SVPSVESL-----HOEVSVLVDPDTVSCLDPP----- 254
Db 656 AEISFNARKVDAYSHMESSGSSRQGHDPDASVTAG-TVVCCLASSLDTSVSASKGSALS 714
Qy 255 -----SHLPD-----QLEDTPILSEDSLEPNFSLAPFVSGGLYGI 290
Db 715 GASSPLYPLEVSFLPEAGLAVQGPGLKSLNLSPTPPSKGAPVSTGAPPSPKGAPIVPT 774
Qy 291 DDT-----ELMGAEKLP-LXDSPIVLSALDCPSLNNAATFSLADDSDTS---TSIF 338
Db 775 ESSISSQVPAIILPSPQKTPFVTSRLISAVQSPKVD-----PMSDVTSPKKTAT 829
Qy 339 ASP--TSPPVLGESVLQDNSFDLNGSDAEQEEMETQSSDPPPS---LTQAPDQSTI 392
Db 830 AVPKDTSATLSKSVPAVTSLSPPKAPVAPSENEATIVPTETPTSLKXNALAAATPKETLAT 889
Qy 393 QLHPATSPAVSPTTSP-AVSLVVSPPAASPEISPEVCFAASTVVSFAVF----- 439
Db 890 SIPKVTSP--SPQKTPKSVLKGAPAMTSKKATEI--AASKDVSPSQPKEVPLLQHVPP 945
Qy 440 -----SVVSPASSAVL-----PAVSLEVLDTASVTSFKASP 470
Db 946 TSPPKSPVSDTLGALTSPPPKGPPATIAETPTPKSKPAAASKKTATPSPSEGVATVP 1005
Qy 471 VTSFAA--APTASPAKOVSSFLE--TTADVEEITGEGLTASGSDVMRRRTATPEE 524
Db 1006 LEIPCKKAPKTAAPKESSTSSSKKAPKTAVSKETPSKGV-----AVPLE 1053
Qy 525 VRLPLQHGWR-----REVRIKKGSHRWQ--ETWYTGPCGKRMKQFPEVI---KYLNRN 573
Db 1054 ISLPLKETSATPGEKSASSPKSPKTAGPKET---PPGVTAVPEISLPPKETPN 1109
Qy 574 LV--HSVRRHFSRPMVGVDFFEEDRTPEGLQWQLSABEISRIQ----- 619
Db 1110 ATPNESLAASQKRSKPTSV-----PKETPPG--GVAMPLEIPSAPQAPKTAVPKQIP 1162
Qy 620 ---AITKGRPRNTEKAKTKEVPK-----VKRGRGPPKVKI 654
Db 1163 TPDAVTILAGSPLSPKASKATAAPKAPATPSVGVIASGEISPSPKKTSKTAAPKENS 1222

Qy	655	TELLNKTNRPLKLEAGETLINEEDKAKIAKSKKKQKQVORGECLTTIOQARNKQKE	714
Db	1223	ATLPPKRPKTAAPKRETPATSEGVTAVPSEISPSPTPAGKGVFVTLTPKGAPNALAES	1282
Qy	715	TKSLKHKAKKSAEKGKTKQEKLEKVKREKVKVKEEVTKAKPACKADTKL	774
Db	1283	PASPK---KVPKTAAPETSTTPSPQKTPKVGPK-----EASATPPSKKTKPT-	1328
Qy	775	ATQRLEERQKQMLLEBMKPTEDMCLTDHQPLPDFSRVPLGLTI-----PSGAFSD	826
Db	1329	-----AVPKETSPSEGVTAVPLEIPSPKAPKTAAPKETPAPSEGATTA	1375
Qy	827	CLTIVFHLHSGKVLGFDPAKDVPSLQVLQGLLCCQGSLSGEVODLLVRLILKAALHDPGF	886
Db	1376	PVQIPPSPRKSKAGSKETPTPS---PEGVTA-----APLEIP--	1412
Qy	887	PSYCOSLKILGEKVEIEPLTRDN-VSEILRCFLMAYGVKPALCDRLRTPQFAOPPOOKA	945
Db	1413	-----ISSKTKTSKNASPKETLVTTPSSKKLSQTVGPKETSLLEGATAVPLEIPPSHKKA	1464
Qy	946	AVLAPFVH-ELNGSTLIINEIDKTL-ESMSSVRKNKWIVEGRLLRLKTVLAKTRGRSEVE	1003
Db	1465	PKTVDPKQVUTPSP---KDAPTTLAESPSFKKAPKTAAPPSERVTVPPEKP-----	1515
Qy	1004	MGRPECLGRRRSS-RIMEETSGMEEEBESIA-AVPGRRGRD-----GEYDA	1051
Db	1516	-ATQKASGTTASKVPVPAETQEVAVSSRETPTVPAVPPVKVNPSSHKTKTIELKEAPA	1574
Qy	1052	TASSIPELEROIEKLSKQQLPFRKLLHSSOMLRAVSLGQDRYRRRYWVLYLAGIFVEG	1111
Db	1575	TLPPSPKSPKIPSSKAPRTSAPKEFPASPSPKPVTTSS-----LAQTAPPSLQK	1624
Qy	1112	TEGNLVPEE-----VIKETDSLKVAHASLNPALFS-----	1143
Db	1625	APSTTIPKENLAAPAVLPVSSKSPAAPARASASLSPATAAPQTAPKEATTIPSCKKAAT	1684
Qy	1144	-----MKMELAGSNTTASSPARASRPLKTKPGFMQPRHFK	1179
Db	1685	ETPIETSTAPLEGAPKETSETSVSKVLMSSPPKASSKKAESTLPTATLPS-LKEASVL	1743
Qy	1180	SPV---RQDSEQPAQLQBAQLHVPAQPOFQLOLQSHKG---FLEQSGSPLSLQGS	1233
Db	1744	SPTATSSGKDS-----HISPVSDACSTGTTTPQASEKLPSKKGPTAFTEMLAFA-	1793
Qy	1234	QHDLSQSAFLSWLSQTSQSHSLSSVLTPSS-----PGKLDPAPOPPPEPPDE	1285
Db	1794	--PESALAITAPIOKSPGANSASSPKCPDFSSKKDKTGLPSAVALAPQTVPEKDTSK	1851
Qy	1286	-----AESSPDLOAFWENISAQMPCNAAPTTPPLAVSDDQPTPSPOQLASSKPMNRP	1336
Db	1852	AIETLLVSPAKGSDCLH-----SPKGVGSGQVATPLAAFTSDKVPPEAVSASVAPKPAP	1905
Qy	1337	SAANPCSPVOFSSTPLAGLAPKR--RAGDPGEMFQSPGTGLQPKRRGRPPPSKFFKQMEQR	1394
Db	1906	AASLTLAP-----SPVAPLPKQPLLESAPGSVLESFSLPVPAEEDLPL-----	1952
Qy	1395	YLTQITAQVPPEMCSG	1411
Db	1953	-----IPPEAVSG	1960
RESULT 7			
QRMSP1			
microtubule-associated protein MAP1B - mouse			
N:Alternate names: microtubule-associated protein MAP1(X); microtubule-associated protei			
C:Species: Mus musculus (house mouse)			
C>Date: 30-Sep-1991 #sequence revision 30-Sep-1991 #text_change 09-Jul-2004			
C:Accession: S07549; S44387; A33645			
R:Noble, M.; Lewis, S.A.; Cowan, N.J.			
J. Cell Biol. 109, 3367-3376, 1989			
A:Title: The microtubule binding domain of microtubule-associated protein MAP1B contains			
A:Reference number: A33645; MUID:90094539; PMID:2480963			
A:Accession: S07549			

A:Molecule type: mRNA			
A:Residues: 1-2464 <NOB>			
A:Cross-references: UNIPROT:P14873; EMBL:X51396; NID:G52999; PIDN:CAA35761.1; PID:G53000			
R:Sanchez, C.; Padilla, R.; Paciucci, R.; Zabala, J.C.; Avila, J.			
Arch. Biochem. Biophys. 310, 428-432, 1994			
A:Title: Binding of heat-shock protein 70 (hsp70) to tubulin.			
A:Reference number: S44387; MUID:94234720; PMID:8179328			
A:Accession: S44387			
A>Status: preliminary			
A:Molecule type: protein			
A:Residues: 653-663, IC<SAN>			
C:Superfamily: microtubule-associated protein MAP1B			
C:Keywords: microtubule binding; phosphoprotein; tandem repeat			
F:589-786/Domain: microtubule binding #status experimental <MTB>			
F:589-592, 639-642, 649-652, 655-658, 660-663, 668-671, 674-677, 679-682, 683-686, 687-690, 691-69			
R:K-E/D-X)			
F:1861-2064/Region: 17-residue repeats			
F:91,116,351,888,1124,1153,1168,1208,1662,1877,1918,2003,2030,2054,2083/Binding site: ph			
F:147,969,1336,1562,1563,1702,1708,1990,2057,2063,2419/Binding site: phosphate (Thr) (co			
F:1953/Binding site: phosphate (Tyr) (covalent) #status predicted			
Query Match 2.9%; Score 286; DB 1; Length 2464;			
Best Local Similarity 19.0%; Pred. No. 6.3e-05;			
Matches 319; Conservative 250; Mismatches 607; Indels 504; Gaps 74;			
Qy	164	NFDSQELYDSFPDQNFEEVCGIHDPDAAEKEMTSVVAENGTLVCSLELEEXQPELMKC	223
Db	156	SFSQNFIEFTDQIEIGELLSTHP--ANKASLFLCFEGDGWKNLDRHNLQDFINIK	213
Qy	224	GYNGSV-PSVESLHQEVSVLVDPDPTVSCLDPSHLPLQLEDPTILSEDSLE---PFNSLA	279
Db	214	LNGASILPEMEGLS-----EFTEYLSE-SVEVPSPFIDILE	247
Qy	280	PEPVSGG-----LYGIDDTLM--GAEDK-----LPLXDS	307
Db	248	P-PTSGGFLKSLKPCCYIFPGRGDSALFAVNGFMNLINGGSEKSCFWKLIRHLDRVDS	306
Qy	308	PVISAL---DCPSLNNATAFSLADDSDQTSIPASPTSPVLGESVLQDNSFDLNGSD	364
Db	307	ILLTHIGDNLPGIN-----SMLQRKIAEL-----E	332
Qy	365	AEQEMETQSDDFPSPSLTQP-----APDQSTQILHPATSPAVSPPTSPAVSL	412
Db	333	EERSQGSTNSDMWNKLIISPLGVFLVFLNVLNPKDPENIKMKESIEEA-----CFTL	385
Qy	413	VVSPAASPEISPEVCPAASTVVSAPVS-----VVS PASSA-----	448
Db	386	QYLNKLSMKPEPLFRSVGNTIIEPVILFQKMGVGVKLEMYVLNLPVKSSEKEMQVFMQQTGTN	445
Qy	449	-----VLP-AVSLVPLTASVTSPKASPVTSAPAAFPPTSPANKDVSSFLETTADVEEI	501
Db	446	KDKAELLPLNGQEVDIPI-SYLTSVSSLI VNHPP-----ANPAEKIRVLVFPNGSTQVNI	498
Qy	502	TGEGLTASGSDVMRRRIATPEEVRLPLQHGWRREVRIKKGSHRWQGETWYVYCGGRMK	561
Db	499	L-EGLEKLKHLDFLKQPLATQKLTGQVPPVVKQVLKQRA-----DSRES	544
Qy	562	QFPEVVIKYLNRNLVHVSRRRHFSFSPRMPVGDPEERBDTPTEGLQWVLSAEEIPSRIQAI	621
Db	545	LKPATKPVASK-----SVRKE-----SKEETPE-----	567
Qy	622	TGKGRPRNTEKATKEVPKVKRGR-----GRPPKVKITELLNKTNDNRPLKLEAQETLINE	677
Db	568	VTKTSQVEKTPKVESKVLVKKDKPVKTESKPSVTEKVESSEQSPV-KAEVAEQAT	626
Qy	678	EDKAKIAKSKKQKQVORGECLTTIOQARNKKQKQKSLKHKEAKKSAEKEKGKTK	737
Db	627	ESKPKVTK-----DKVVKEIKTLEKEEKPKKEVKKEDKTPLKQDE-----KPR	674
Qy	738	QEKLEKVKREKKEKVKMKKEEYVTKAKPACKADTKLTATQRRLEERQKQMLLEMKKPT	797
Db	675	KEEVKEIKKEIKKEERKELKKEVKKETPLKDAKKEVKKEKKEKVEKPEKKEIKKIS	734

Db 494 EDSRRFSQPSTSHHETSHSPKDSBAYQGRSFPLNYQSOS-----PGHEFLSKSEIK 548
Qy 184 SGIHPD-----EAAEKMTSVVAENGTGLVCSLEEXQPELKMCGYNGSVFVSLEHDEV 239
Db 549 QEFSTTSSASSDLELWEMPON--PLTRMLERHMRFFIDVSVFNRIDEIVELNQKA 606
Qy 240 SV-----LVPPDVTSCLDPP-----SHLPDQLEDTPILSEDS-L 272
Db 607 RASYEKEFTGRFPFKCNNDVLSIQKIVFHPBRDYYVYENPCSELEVRIRDWRKLSDTADL 666
Qy 273 EPFNSL-----APEPVGGGLV-----IDDT-----ELMGAEDKL-----PLXD 306
Db 667 DDFRATDSKELGRDQACGRTSGRPSLDESRTNRLSFDSTHHPAORAORSHSLCIGMTP 726
Qy 307 SPVISALDCPSLNNAFASLLADDSQTST--IFASPTS--PPVLGESYLDQNSFDLNNGS 363
Db 727 STPPPTSQ-PLLVNTH--LPGTSQPSTSGGITTFRSSQPPPLMSPVSRHNSMS-STGR 781
Qy 364 DAQEEMETQSDPFPPLTOP-----APQSSTIQHPATSPAVSPTT 406
Db 782 PASIQTLRHQSVMPFPDVSIPPPPIPTHDEMMAPRGTPPGRSRSETMVLRSF--PFG 838
Qy 407 SPAVSLVVS PAASPEISPEVCFAASTVVS PAVFSVVS PASSAVLP PAVSLEVLPLTASVTSP 466
Db 839 TPIQMLLTMPIVPP--PHLIAATST--GTHSVSSSAHSTPHRSIS-GTPVHCEPNS 890
Qy 467 KASPVTSAAAFTASPANKDVSSPLETTADVVEITGEGLTASGSDVMRRRIATPEVR 526
Db 891 KTSQPPTPK-----GP-----SRPEKQV 906
Qy 527 LPLOHGWREVRIRKKGSHRWQETWYQPCGKRMKQFPEVIKYLNRNLVHVRREHFSFS 586
Db 907 I-----RHDITSKS-----GN-----SNAINALQAR----S 928
Qy 587 PRMPVGDFEERDPEGLQWVLSAEEIPSR-----QAITGKGRPRNTEKAKTEVPVKV 643
Db 929 QSMTSGD---PKKSAPSTFVVRDAGSDLVQAQIMSQPNLGRKUPRIEKKSSALQ-----N 981
Qy 644 RGRGRPPVKITELNKTNDNPLKLEAQETLNEEDKAKIAKSKKMKQKQVORGECLTTI 703
Db 982 IQNHQPHSN---ANSTPSTPSTHQAMFKDKERKKEKKEERERARERMKRKE 1037
Qy 704 QGOARNRKQBTSLKHKAAXKAEKGTQKQKLEKVKYREKKEKVKMKKEBEVTK 763
Db 1038 TKBERNKKEKEMERAKLEDERQERKREKKEKDERKKEKV-RKKAEEKELKKKKH-RK 1095
Qy 764 AKPACKADKTLATQRRLEERQKQOMILEMKKPTEDMCLTDHQPDPFSRVPGLTLPSCA 823
Db 1096 GDSDESDDSDNDELDDVRKSTKEMTQEEK-----DHQ-----LALLSKGG 1138
Qy 824 FSDCLTIVEFL-----HSFGKVLGFPDPAKDVPVSLVLOBGLLCOGD-----SLGE 868
Db 1139 -----IINLKSRRSRDKRAHDSPEKMQQKQSRVLIESDDEGGK-DGDKGNSSNGE 1191
Qy 869 VODLLVRLKLAALHDGPP-----SYCQSLKILGEK-VSEIPLTRD-----NVSEILRCF 917
Db 1192 ESD-----SEKADLPPPPAPPSLSSESADQLKVLKEREKELTSSDDSDHNDAGEIHQ-- 1245
Qy 918 LMAYGVXPALCDRLATQTFQAPPOQKAAVLAFPVHELNGSTLINEIDKTLSEMSY-- 975
Db 1246 -----QRLTED-----RENRKQKSLTAYSS 1266
Qy 976 -----RKNKVIWEGRLRR--LKTVLAKRTG-----RSEVMGPECLGRRRSR 1018
Db 1267 DEQGERKN---VPKMRRRDDSEDAAKHPGWSAKDDQQRKRUKEHRRSSDESKKNKR 1323
Qy 1019 IMEETSGMEEBEEESIAAVPRGRGRDGEVDATASSIPELERQTEKLKSRQL----- 1071
Db 1324 DFRDIPHDVDSDEET---EDGSRSRQ-STSSSTISNVTAKERK-EKSGKTPLRVPEPT 1378
Qy 1072 ---FFRKLLHSSQMLRAVSLGQDRYRRYVWLPVLAGIFVEGTGNLVPBEVIKKTDS 1128
Db 1379 GTPLLSPLKILSPKHLSPKSTSTSKRSS-----ISDHE-NLISPRQNRITTS 1425

Qy 1129 LKVAHAASLNAL-----FSMKOMELAGSNTTASSPAR 1160
Db 1426 TSTATTSKHEALGPEKPLSPPTAKSSVSSIDDPSIRDEFSM-----NSAADSPMS 1478
Qy 1161 ARSPL-----KTKP-----GFM-----QPRHFKS----- 1180
Db 1479 TTGRPVLTKAAMKAFNSTPPKKVSYSLIDCYMLGMMAKNSSGQHDSGSSSDSS 1538
Qy 1181 -----PVRQDSEQO----- 1191
Db 1539 DGSTSSDDSDDEVPKQTEPVTSPVVASDNGSPENVVETPSIVSQTPREPEPTISEQ 1598
Qy 1192 -AQLOPEALHVPAPQPOLQLOQSHKGF-----LEOEGSPLSLG 1231
Db 1599 SSESPEBA---VPECPEASVEPQMETSONVEPVESEHEDSHHEHGDSEVAVESQQPLEHQ 1655
Qy 1232 QSOHDLQS-----AFLSWLSQTSQSHSLSSSVLTDPSSPG-----KL 1270
Db 1656 EEKBELENKILDVAAEHHEEQVQDEDSVSSIIPAPDEPDVTQAEKSAHTLISDOET 1715
Qy 1271 DPAPSQPPEPEPEDEAESPD--LQAFWFNLSAQWPCNAAPTPL----- 1313
Db 1716 DQAVQSIFDEEAEDEFQYDFGISTNEKEVSGKDPHNKTEPLNNGHTDLLFSPSSA 1775
Qy 1314 AVSEDDQTPS-----PQULASSKPMNRPSAANPC-SPVQFSSTPLAGL 1355
Db 1776 HASEKQSTKSDDEEDSELVMEKEVPMEQVIAQVHVHVPSEPSMEBEVKLETSPV-- 1832
Qy 1356 APKRAGDPGEMPOSPGTLGQPKRGRPPSKFFQKMEQRYLTQLTAQVPPEMCSGWWMI 1415
Db 1833 -PKE---EPKMEESP-----EQ----- 1847
Qy 1416 PDPEMLDAMLKALHPRGIREKALKHLNKHKRDLEQVCLRPSADPIPEPROL-PAFOEGI 1474
Db 1848 PTPOLISNSESQDTP-----GAVNNHLHENHDAVQ-----TPIQLOPASQHV 1890
Qy 1475 MSWSPKEXKTYETDLAVLQWVEELEQRVIMSLOLIRG-----WTCPSDPSTRE-DLAYCE 1527
Db 1891 AQPSPRP-----AVAPDSQQNGPVLVSQSQSPSPMSQQSDMA--Q 1929
Qy 1528 HLSDSQEDI 1536
Db 1930 NLILSKDI 1938

RESULT 14

A44265

trithorax homolog HTX, version 2 - human

N:Alternate names: acute lymphoblastic leukemia gene 1 protein (ALL-1)

C:Species: Homo sapiens (man)

C>Date: 30-Apr-1993 #sequence revision 18-Nov-1994 #text_change 09-Jul-2004

C:Accession: A44265; A44264; I58112; I37165; I38485

R:Kachuk, D.C.; Kohler, S.; Cleary, M.L.

Cell 71, 691-700, 1992

A:Title: Involvement of a homolog of Drosophila trithorax by 11q23 chromosomal translocation

A:Reference number: A44265; MUID:93046667; PMID:1423624

A:Accession: A44265

A:Status: nucleic acid sequence not shown; not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 1-3968 <TKA>

A:Cross-references: UNIPROT:Q03164; GB:L04284; NID:G184393; PIDN:AAA58669.1; PID:G184394

A>Note: sequence extracted from NCBI backbone (NCBI:P:117729)

R:Gu, Y.; Nakamura, T.; Alder, H.; Prasad, R.; Canaani, O.; Cimino, G.; Croce, C.M.; Can

Cell 71, 701-708, 1992

A:Title: The t(4;11) chromosome translocation of human acute leukemias fuses the ALL-1

A:Reference number: A44264; MUID:93046668; PMID:1423625

A:Accession: A44264

A:Status: not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 63-316, 'GLLINSELEK', 327, 'Q', 329, 'VR', 332, 'DKEGTTP', 340, 'T', 342, 'EDKTVVRQSPRI

546, 'LQIESTSP', 2555-3554, 'N', 3556-3594, 'V', 3596-3899, 'A', 3901-3968 <GUL>

A:Cross-references: GB:L04731; NID:G339921

A>Note: sequence extracted from NCBI backbone (NCBI:117779)
R:Diabali, M.; Selli, L.; Parry, P.; Bower, M.; Young, B.D.; Evans, G.A.
Nature Genet. 2, 113-118, 1992
A:Title: A trichorax-like gene is interrupted by chromosome 11q23 translocations in acute
leukemia
A:Reference number: 158112; MUID:93265134; PMID:1303259
A:Accession: 158112
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: mRNA
A:Residues: 1317-1700, 'DD', 1703-1936, 'H', 1938-2180, 'S', 2182-2328, 'DIA'
A:Cross-references: GB:L01986; NID:9307522; PIDN:AA92511.1; PID:9553800
R:Marchalek, R.; Grell, J.; Lochner, K.; Nilsson, I.; Siegler, G.; Zweckbronner, I.; Bec
Br. J. Haematol. 90, 308-320, 1995
A:Title: Molecular analysis of the chromosomal breakpoint and fusion transcripts in the
leukemia
A:Reference number: 137165; MUID:95315013; PMID:7794749
A:Accession: 137165
A:Status: translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1212-1603, 'GTE', 'MAP'
A:Cross-references: EMBL:X83604; NID:9897757; PIDN:CAAS8584.1; PID:9899268
A>Note: submitted to the EMBL/GenBank/DBJ databases by R. Marchalek, 20 December 1994
R:Gu, Y.; Alder, H.; Nakamura, T.; Schichman, S.A.; Prasad, R.; Canani, O.; Saito, H.;
Cancer Res. 54, 2327-2330, 1994
A:Title: Sequence analysis of the breakpoint cluster region in the ALL-1 gene involved in
leukemia
A:Reference number: 138485
A:Accession: 138485
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1251-1486, 'G', 1488-1538, 'RES'
A:Cross-references: EMBL:U04737; NID:9451554; PID:9451555
C:Genetics:
A:Gene: GDB:MLL; HTX; ALL-1; HRX
A:Cross-references: GDB:128819; OMIM:159555
A:Map position: 11q23-11q23
A:Introns: 1338/1; 1362/3; 1406/3; 1444/3; 1493/3; 1525/3; 1566/1
A>Note: the list of introns is incomplete
C:Superfamily: histone methyltransferase, trithorax protein type
C:Keywords: acute lymphoblastic leukemia; proto-oncogene; zinc finger
F:1434-1456/Region: zinc finger CCHC motif
F:1479-1506/Region: zinc finger CCHC motif
F:1527-1556/Region: zinc finger CCHC motif
F:1569-1596/Region: zinc finger CCHC motif
F:1873-1900/Region: zinc finger CCHC motif
F:1933-1955/Region: zinc finger CCHC motif

Query Match 2.8%; Score 272; DB 2; Length 3968;
Best Local Similarity 17.8%; Pred. No. 0.00049;
Matches 398; Conservative 247; Mismatches 707; Indels 878; Gaps 99;
Qy 17 LPPAP-...AASGLKPS-...SSGGLYTGSPMFPQQGKSLNGDVNVL 61
Db 40 LPPGPPVGGGFGAPPAPVAAAAAAGSGAGV-...PGGA 78
Qy 62 STVSHTTTGILNAPHSSSTSH---LHPSVAYDCLMNYQYPSANPGSNLKDPLLSQ 118
Db 79 AAASAGSSSSSSSSSSSSSAGSALLRVGPGDA---ALQVSAAGTTLNRFRVAFV 134
Qy 119 FSGQYPLNGILGS-...RPS-SPSHNTN-...LRAGSKFW 153
Db 135 ESGGG-...GGSGEDQFLGFGSDEVRVRSPTSPVKTSPRGRPRSGSDNS 187
Qy 154 A---NGTHSPWGLN-...FDSQEL-...YDSPDQ 177
Db 188 AILSDPSVFSPLNKSEKSGDKIKKOSKSTKRGRRPPTPGVKIKITHGKDISLPGK 247
Qy 178 NFEVCSGIHPDEAAEKEMTSVAENGTLGYCSL-...ELEEXOPELKMGYNGSV 230
Db 248 NKEDSLKKIKRTPSATFOATKIKLRAGKLSPLSKFKTKGLQIGRGVQIVRRGRPP 307
Qy 231 SVESLH-QEVSIV-...PPTVSCLDLDPDQLDTEPILSEDSLEPPNSLA-... 279
Db 308 STERIKTPSVSSILNWKSPKSGKTRKEHLHLQKIRQ---LSDKALEGLSGLFLQ 364
Qy 280 ---PEPVSGGLY-...GIDDTLMGAEDKPLXD-...SPVISALDCPSLNN 320

Db 365 KGQWQPLLNSYRQKKGAKKIEKAAQKGRVKVQVKVIRQFIMPVVAISSRIKT 424
Qy 321 ATAFSLIADD-...SQTSTSFAPSTPPVVLGSEVLQDNFDLNGSD-ABQE 368
Db 425 PRRF--IEDEDYDPPFIARLESTPNRFRFAPSGSSEKSAASQHSQMSDSRSSSP 482
Qy 369 EMEQTSSDFPPLTQAPDQ--SSTIQLHPATSPAVSP-...T 405
Db 483 SVDTSSTQSASEEIQVPEKSSSTPEVHPPLPISQSPENESNDRRRSRYVSERSFGRT 542
Qy 406 TSPAVSLVWSPAAPEISPEVCPAASTVSPAVSVSPASSAVLPAVSLSEVPLTASVTS 465
Db 543 TKKLSTLQSAPOQQTSSPP--PPLLTPPP-...LQPASS-...ISDHTPMLMPPTI 590
Qy 466 PKASPVTSAPAAFPPTASPANKDVSSFLETT-...ADVEETGGGLTASG 510
Db 591 PLASFFL-PAS--TAPMQGKRKSLREPTFRWTSKLHRSRSEPOYFSSAKYAKGLIRKP 646
Qy 511 SGDVRRRIATPREV-...RL--PLQHWRRVRIRKKGSHRMOGETWY 553
Db 647 IFDNFRPPLTPEDVGFASGFSAGTAAARLPSLHSGTRFDWH-... 691
Qy 554 GPCGKRMKQPEVIKYLRLVSVRREHFSFSPMPVGDFFERDTPGELQWVQLSABE 613
Db 692 ---KRSP-...LLRAPR-...FTPEAHSRIFESVTLPSNRTSAGTSSSG 730
Qy 614 IPSRIQAITGKGR-...PRNTEKAKTKEVPKVRGR-...GRPPKV-... 653
Db 731 VSNR-...KRKRVFSPISRSPSPSHSMRTSGRLSSSELSPPLTPPSSVSSLSISV 784
Qy 654 ---ITELLNKTDRPLKLEAQETLNEEDKAKIAKSKKMR-... 691
Db 785 SPLATSAINTFTFPSSHSL-...TQSGESAENQRPKQTSAPAEPTSSSPTPLFPWPT 840
Qy 692 --QKVRQGECLTITQGOARNRKOETSLKHKEAKKXAEKEGKTKQEKLEKVKREK 749
Db 841 PGSQTERG-...RNKAKAPELSKORDADK--SVEKDKSRERDRERENKRES 888
Qy 750 KKKVMEKE-...EVTAKAPACKADKTATQRLERQKQOOLMEEMKQTE 798
Db 889 RKEKKGSEITQSSSALYPVGRVSKVGVEDVATSSAKKATGRKSS-... 937
Qy 799 DMCLTDHQLPDSRVPGLTLPSCAFSDCLTI-VEFLHSFGK-...VLGDPKADVPSLG 853
Db 938 ---SHDSGTITSVTLGDTTAVTKILIKKGRNLEKTNLDLGTAPSL 984
Qy 854 VLQEGLLCOGDSLGEVQDLLVRLKALHDPGFPSPYQCSLILGEKVS-...IFLTRDNV 910
Db 985 --KEKTL-...LSTPSSSTVKHSTSSIGSMLAQADKLPMTDKRV 1023
Qy 911 SEILRCFLMAYGVXPALCDRLTQPPQAPQKAAVLAPFVHELNGSTLIINEIDKTL 970
Db 1024 ASLLK-... 1032
Qy 971 SMSYRNKNIWEGRLRLKTVLAKTGRSEVENRPEECLGRRRSRIMEETSGMBEE 1030
Db 1033 QLCKIEKSL-...KOTDQPAQ-...GQSDS 1058
Qy 1031 EESIAAVPGRGRGRDGEVDATASSIPELEROIEKLSKQLFFRKLHLLSHQMLRAVSLG 1090
Db 1059 SETSV-...RGR-...IKHVCR-...AAVALG 1080
Qy 1091 QDRYRRYRWLPLYLAGIFVETEGENLVEEVIKKTDSLVKVAHASLNALPFSM-...KM 1146
Db 1081 RKR-...AVFPDDMTLSALPWEEREKILSSMGNDK 1114
Qy 1147 ELAGSNTTASPARSRPLKTKPFMOPRHFSPV-RGQDSEQ-...PQALQPEALH 1201
Db 1115 SIAGSD--AEPLAPPIKPKVTRNKAPQ--EPVVKGRSRRCGCGPCQVEDCCVC 1170
Qy 1202 VPAQPPQL-...QLQLOSHKFLBQEGSPLSLGOSQHDLQSLQSAFLSW 1245

Db 1171 TNCIDKPKFGRNIIKQCCMKRKNQNLQWMPKAVLQKQAKVKKKKKSTSE----- 1224

Qy 1246 LSQTQSHSLLSSSLVLTDPSSP---GKLDPAQSPQ-----PPEPEPDEAESPPDLQAFWF 1297

Db 1225 ---KDSKSSVKNVVDSSQKTPPSAREDPAPKSSSPPPKPKPVEEKSEEG----- 1274

Qy 1298 NISAQPCNAATPTPLAVSDDOPTSPQOLASSKPMWPSAANPCSPVQFSTPLAGLAP 1357

Db 1275 NVSAPGP-----ESQATTPASRKSSKQVOPALVIPPQP-----PTTG--P 1314

Qy 1358 KRRAGDPGBMPQSPGLGQPKRGRPPSKFFKQMEQRVLTQLTAQPVPPCMSCGWWIIPD 1417

Db 1315 PRK-----EVPK--TTPSEPKK-----QPPPE--SG----- 1338

Qy 1418 PEMDLAKLHPGRIKALKHLNKHDFLOEYCLRPSADPIPEPQLPAFQEGIMSW 1477

Db 1339 ---PEQSKQK-----KVAPRS-----IPVKQK----- 1358

Qy 1478 SPKEITYETDLAVLQWVELEQRVIMSDLIQRTGWTCPSPDSTREDLAYCEHLSDSQEDIT 1537

Db 1359 ---PKEKEK-----PPVKNKOENAGTNILSTLS----- 1384

Qy 1538 WRGPGREGIAPORTKTNPLDLAVMLAALQONVRYRLREPLWPTHEVVLEKALLSTENG 1597

Db 1385 ---NGSSKQKI--PAD-GVHRIRV---DFKEDCEAENWEMGGIGI---LTSVP-- 1427

Qy 1598 APEGTTTISVEITPRIIRWOTLQRCRSAHV-----C-----LCLGHLERSIAWE 1644

Db 1428 ---ITPRVVCF-----LCASSGHVEFYQVCCPEPHKFCLENERPL--E 1468

Qy 1645 KSNVKTCLVC-----RKGNDLFLLLDCGDRGCHYCHRPKWEAVP---EGDWFT 1694

Db 1469 DQLENWCCRCKFCHVCRQHOATKQLECNKCRNSYHPECLGNPNYTKPTKKKVMICT 1528

Qy 1695 VCLAQVGEFTQKFPKRGQKSGYSLNFSEGDGRRRVLKGRSPAAGPRYSER 1754

Db 1529 KCVRCCKSGSTT-----PGKGWDAQWHDLSLCHDCA---KLFAKGNFCPLCDKCYDDDD 1580

Qy 1755 LSPSKRRRLSM--RNHSDLTFCETILMEM-----ESHDAW----- 1789

Db 1581 YE-SKWMQCGKDRVHVK---CENLSDMEYELSNLPSVAYTCVCTERHPAEWRLAL 1636

Qy 1790 ---EKELQISLQVLTALLNSRTTSHLLRYQAAKPPDLNPETEESIPRSRSPGPPVLT 1696

Db 1637 GKELQISLQVLTALLNSRTTSHLLRYQAAKPPDLNPETEESIPRSRSPGPPVLT 1696

Qy 1802 GYRIIKNPMDESTRERLLRGVTSSEFAADALLVFDNCOTFNEDDSEVGKACHIMR 1861

Db 1697 VSKQKQQLLEGVYKRWQDQNTYSVLEFSDDIVKIIQAAINSQGGQPEIKKANSWYKS 1756

Qy 1862 PFESRWEFFY 1871

Db 1757 FFIRQWVVF 1766

RESULT 15

T03455

ALR protein - human

C:Species: Homo sapiens (man)

C:Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 09-Jul-2004

C:Accession: T03455

R:Prasad, R.; Zhadanov, A.B.; Sedkov, Y.; Bullrich, F.; Druck, T.; Rallapalli, R.; Yano, Oncogene 15, 549-560, 1997

A:Title: Structure and expression pattern of human ALR, a novel gene with strong homolog A:Reference number: Z14954; MUID:97388474; PMID:9247308

A:Accession: T03455

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-4957 <PRA>

A:Cross-references: UNIPROT:O14686; EMBL:AF010404; NID:g2358286; PIDN:AAC51735.1; PID:g2

C:Genetics:

A:Gene: ALR

A:Map position: 12

C:Superfamily: acute lymphoblastic leukemia protein, ALR type

C:Keywords: alternative splicing

Query Match 2.6%; Score 258.5; DB 2; Length 4957;

Best Local Similarity 18.6%; Pred. No. 0.0026;

Matches 365; Conservative 255; Mismatches 670; Indels 673; Gaps 92;

Qy 15 TGLP-----PAPAASG-----LKPSPSSGGLYTNQSPMNFPOQG----- 49

Db 2279 TRLPGPGEPVGPAGPAQFIELRNVQKGLGPGGTP--FPGQGPQRPFRFVPSVSDPHRL 2336

Qy 50 ---KSLGNDVNVGLSTVSHTTTSGILNSAPHSSSTSHLHPSVAYDCLWNYSQYPSANPG 107

Db 2337 APEGLRG-LAVSGLPPQK-----PSAPAPALNNSLHPT-----PHTKGTPLTPTG 2380

Qy 108 SNLKDPPLLSFGSQVPLNGILGSRQSPSPSHNTNLRAGS----- 149

Db 2381 LELVNRPPSSTELRPNPL-ALEAGKLPCEDEPDLDDDFDAKHALEDDELAHLGLGVDA 2439

Qy 150 -----QKFWANGTHSPMGLNFDQELY-----DSF-----PD 176

Db 2440 KGDDDELGTLENLTDNPHLDDLLNGDEFDLLAYTDPDLTDGKKDI FNEHLRLVESANEE 2499

Qy 177 QNFEEVCSGIHPD-----EAAEKMTSVVAENGTLVCSLEBEXQPLKMG 224

Db 2500 AEREALLRGVPGPLGPEERPPPAADASEPLASVLPE-----VKPVEEGRRHPSPCQ 2553

Qy 225 YNGSVPSVESLHQEVSVLVPDPTVSCLD--DPHSLPDQLEDTPILSEDSLEPFS--- 277

Db 2554 FTIATPKVE-----PAPANSGLGLKPKQSGMMGSRDTRM-----GTGPFSSSGHT 2599

Qy 278 -----LAPEPVS--GGLYGIDDTTEL-MGAEKLPFLXDPVISALDCPSL 318

Db 2600 AEKASFGATGPPAHLTTPSLSPGSGSLLEKELESQA--LTLPGPAAASGDDELDM 2656

Qy 319 NNATAFSLADDSTSTSI PASPTSPVGLGESVLDNSFDLNGSD-----AQEWE 371

Db 2657 ES-----SIVASEL-----PLLIEDLLEHEKELQKQQLSAQLQPAQOQOQ 2699

Qy 372 TQSSDFPPLTQAPADOSSTIQLHPATSPVPTTSPAVSLVSPVPAASPEISPEVCP-- 428

Db 2700 QQQHSLIPA---PGPAQMSLP-HEGSSPSLA-GSQQLSLGLAVARQPGIPLMPTQP 2754

Qy 429 ---AASTVSPVSVVSP-----ASSAVLPVSVLEVPLTASVTSKPAVSPVSPA 476

Db 2755 PAHALQRLAPSMAMVNSQGHMLSGHQAGLVPOQS-----SQPVLSSQKPMGTMPFSM 2809

Qy 477 APPTASPAKD--VSSPLETTADVETGEGLTAGSGDVNRRRIATPEEVRLPLQHW 534

Db 2810 CMKPQQLAMQQLANSFPDPT-DLDKFAAEDII---GPIAKAMVALKGIKKVMAQO-- 2862

Qy 535 REVRIKSGSHRWQGETWYVPCGKRMKQFPEVIKYLNRNLVHSVRREHFSFSPMPVGD 594

Db 2863 -SIGVAPGMNRQ-----VSLAQRL-----SGP----- 2886

Qy 595 FEEDTPEGLQWVQLSAEIPSRIOAITGKGRPRNTEKAKTKEVPKVRGRGRPPVKVI 654

Db 2887 -----SSDLQNHVAGSGQ-----ERSAGDP----- 2907

Qy 655 TELINKTDNRPLKLEAQETNEEDKAKIAKSKKMKQKQVORGECITTIQOARNRKQE 714

Db 2908 -----SQPRNPPTFAQGVINEADQRYEWEFLHTQQLQWQ--LKVLEBQIGVHRKSR 2959

Qy 715 TKSL--KHKEAKKSAEKGKTKQEKLEKVKREKKEKVKMEKEVETKAKPACKADK 772

Db 2960 -KALCAKORTAK--AGREPEADAEKL--KLVTEQOOSKIQ-KOLDQVRKQ---KEHT 3009

Qy 773 TLATORLEERQKQOMILEEMKKTEDMCLTDHPLDFSRVPLTLP-----G 822

Db 3010 NLMAEYRNKQOQ 3069

Qy 823 AFSCLTIV-----EPL-----HS-----FGKVLGFPDAK--DVPSL 852


```
Qy 787 QMILEE-----MKPTEDMCLTDHQLPDPFSRVPGLTLPSCAFSDCLTIVEFLHSFGKVL 841
Db 1559 ILAMESPATSVRVKPTVEDSTIQ-----TTPKNEAGSSLLV-----EQVQOVL 1601
Qy 842 GFDPKADVPISGLVQLCGDLSIGEVDLLVRLKAAALHDPGPPSYCQSLKILGBKVS 901
Db 1602 QEDDAQASAGTIEDEDYSNQ-QAAAAVENANSQOLDANHYGPNPESKQO-----QORS 1655
Qy 902 BIPUTRONVSE-----ILRCFLMAYGVXPALCDRLRT--OPFOAPPOQKAAVLAPPV 952
Db 1656 KSGSTRPMFSPGTRPPFRIFEFKWSYTHORLLSDVLFSLFTDQVWRSHSTKSVLDF-V 1714
Qy 953 HELNGSTLIINEIDKLTLESMSYKKNWIVGRLRLKTVLAKTG-RSEVEMGRPRECL 1011
Db 1715 NSSNAIFVFN-----TVHLISQLADNLIACGGLLPL-----LASATSPNSELD----- 1759
Qy 1012 GRRSSRIMETSQMEBEESIAAIVGRRGRDGEVDATASSIPELERQIEKLSKRQL 1071
Db 1760 -----VLEPTQGM--PLEVAVSFLQRLVMADVLIFATSLNFGLEAE-KNWSGGI 1808
Qy 1072 FFRKLLHSSQMLRAV-----SIGQDRYRRYVW-----LPYLAGI----- 1107
Db 1809 L-----RQRLVCTCAVRNCKLECKERTRYNVGALARDVPGAHLQALIRGAQASP 1860
Qy 1108 -FVETEGNLP-----EVIKKTDSLKVAHAASLNPALFSKMKELAG 1150
Db 1861 NIVESITGQSPVKDPEKLLQMDMNRRAVIVYDVEETKQAFSLAIVFISVLVMSK 1920
Qy 1151 SNTTASSPAR-----ARSRL-----KTKGFMQPRHFKSPVRQD 1186
Db 1921 YRDILEPAPQIQSPVLQRTAGGNAASARPLFPQWSHHVVYQFLPESHQNHNSNQH 1980
Qy 1187 SEQPQALQPPRAQLHVPAQPOQLQLQSHKG-----FLEQSGSPISLGOSQ----- 1234
Db 1981 QQQQQQQQQQQQQHYQQQQQQQVHSHHMTAAAYQQQQQVAAQQQQSHPTPLA 2040
Qy 1235 -HDLQSQAFSLWLSQTSQSHSLSSSVLTPDSSFGKLDPAPOPEPEPDEARESSP 1290
Db 2041 THSTSSASSTATSQPASSSLSLASQSQSHRLHKQOQQQQQQQQQQQPHYPH 2097

RESULT 18
T34513
hypothetical protein ZK783.1 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 09-Jul-2004
C;Accession: T34513
R;Favellio, A.; Vaudin, M.
submitted to the EMBL Data Library, August 1994
A;Description: The sequence of C. elegans cosmid ZK783.
A;Reference number: Z21536
A;Accession: T34513
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-3507 <Fav>
A;Cross-references: UNIPROT_Q23587; EMBL:U13646; PIDN:AC24418.1; GSPDB:GN000021; CESP:ZK
A;Experimental source: strain Bristol N2; clone ZK783
C;Genetics:
A;Gene: CESP:ZK783.1
A;Map position: 3
A;Introns: 14/1; 48/2; 84/1; 196/3; 303/1; 381/1; 586/1; 605/1; 1175/3; 1207/1; 1409/2;
3504/1

Query Match 2.5%; Score 250.5; DB 2; Length 3507;
Best Local Similarity 18.6%; Pred. No. 0.0039;
Matches 253; Conservative 157; Mismatches 537; Indels 415; Gaps 50;

Qy 208 VCSLELEE---XQBELKMKCG-----YNGSVFSVE-----SLHQEVSVLVPD 245
Db 1820 LCELEKQCTVDEBEVQCCACLPGHHPINGTCQSLQISGLCAQKNDKNKHAECIDIHPD 1879
Qy 246 P-TVSCLD-----DPSHLPQLEDTPILSEDSLEFPNLSLAPEPVSGGLYGIDD 292
```

```
Db 1880 SHFCSDFGFIGDMICDDVDDECNAGMCDDEKTCENTIGSFNCVCL-----GFKKVD 1935
Qy 293 TELMGAEDKULPLKDPSPVISALDCPSLNA-----TAPSLADDSTQTSIIFASPTSPVVLG 348
Db 1936 -KCVDEKQPNREKIBIDEENSSNSGQEKPTTKGIVSSTATSSE---STTAEPHVT 1991
Qy 349 ESVLQDNF-DLANNGSDAEQEEMETQSSDFPPSLIQ-----PAPQSSITQLHPA 397
Db 1992 TSISSSTTSTKDMTSSKSPENVTWSSSEPEVSTSSSKTTTASSETTVSTPSSSESEAPLT 2051
Qy 398 TSPAIVS---TTSPAVSLVSPAAPEI-----SPEV-----CPAASTVSPVAFVS 440
Db 2052 SSPATTEVITESSVKSTTPKESSSEITVKLSKSPVTESSVKSPSPSTTSQSVTS 2111
Qy 441 VVSPASAVLPVLSVPLTASVTSPKASPVTPGAAAPFPASPANKOVDSFLETTADVEE 500
Db 2112 TVPETSASTV--LSSEAPVTS--TSPTEVHTSETKPSLSASSTTGDNTTPTSTSLAS 2167
Qy 501 ITG-----EGLTASG-----SCDMRRRIATPEEVRLPLOHGWREVRKIKGSHRWQ 548
Db 2168 VKTSAPEGTSASVAPVKLSLSPDVSPQSTKTFDATE-----SSTVQA 2211
Qy 549 ETWYVYGGKRMKQFPE---VIKYSRLNLVHSVRRHFSFSPRMPVCDPFPERDTP 604
Db 2212 SETSSGTSVKSTSE-PESHVTKLSITSSNFPSSVPVTSPTKSTPTVPSTEQPTSTTSGQ 2270
Qy 605 QMWQ-----LSABEIPRIQAIQKGRPRNTEKAKTKVDPK-----VKRG 645
Db 2271 SLTPMNSSEVLTTSEPHVLSLSLSPVQSST---TPNNLSESTVETPKTSSESLSN 2326
Qy 646 RGRP-----PKVKITELLNKTDRPLKLEAQETLNEEDKAKIASKKMRQKQVORBECL 700
Db 2327 SEBFTTEAPTTLSPDLSTTTN---NLSQSTSVSTEDRSEI---SSENSEKPTSABELV 2380
Qy 701 TTIQGAQRNKRKQBTSLKHKEAKKSAEKGKTKQEKLKVKRKEKEKVKMKKEKE 760
Db 2381 TS-----SVTHVASSPD-----VPTESSE----- 2400
Qy 761 VTKAKPACAKADMTATORRLEERQKQMILEEMKKPTEDMCLTDHQLPDPFSRVPGLTLP 820
Db 2401 -----PDDLTCGSTENIPEASSKQTI---SSTPTPTTTTASEEPTKSTMSPLDSTT 2449
Qy 821 SGAFSD-----CLTIVFEL---HSFGKVLGDFPAKOVPSLGLVL 855
Db 2450 SNVLSSESTTPSSSKSPVSSSTEGISVVTSTBFKVPVESTISSVLEEDTKTTTPS-PIL 2508
Qy 856 QEGILLCOGDSLGEVDLLVRLKAAALHDPGPPSYCQSLKILGBKVSIPILTRDNVSEILR 915
Db 2509 EETTTASETSEPLTEDSLT-----VSVRIHELTTSENV----- 2542
Qy 916 CFLMAYGVXPALCDRLRTQFOAQPOQKAAVLAPFVHELNGSTLIINEIDKLTLESMSY 975
Db 2543 -----PKESSESTTSSSEKPSQEPAGILTTSTVVVPTSSVSLI-----TASEIEAI 2588
Qy 976 RKNKWIVEGR---LRLKTVLAKTGSEVEMGRPECLGRRSSRIWEETSGMEEEEEE 1032
Db 2589 TSNTFPFQGRTPITTSKPSLVKSTTSPSTVTSSEPSSESTKRTTVSTVTTTPTTEETTS 2648
Qy 1033 ES--IAAVPGRRRGRDGEVDATASSIPELERQIEKLSKRQLFFRKLLHSSQMLRAVSLG 1090
Db 2649 ESILHTAP----- 2657
Qy 1091 QDRYRRRYWVLPYLAGIFVEGTGNLVPBEVIKKTDSLKVAHAASLNPALFSKMKELAG 1150
Db 2658 -----SKPTSTTESSEAPTTPAKTS----- 2678
Qy 1151 SNTTASSPARSRPLTKXGFMQPRHFKSPVRQDSEQOQALQPOEALHVPAPQO--- 1207
Db 2679 -----ETKPSNVSSSTRKSTENVETSTOSGSLSESTMSSTSEPTENA 2722
Qy 1208 PQLOLQSHKHGFLEOBGSPLSLQSQOHDLSQSAFL--SWLSQTSQSHSLSSSVLTPDS 1265
```


Db 2723 PAVTVSSSEASSSTLSEASVSKLSLPPSPITSEAVTVSSRAPAEITMSSE 2782

Qy 1266 SPGLDPPAPQPPEPEDEASSPDLOAFWNISQAQPCNAAPTPPLAVSDDP----- 1320

Db 2783 SHREISTVSSSEPEIPLSITVSPNV-----VTASSTPSEPIILSSVTSSTPRVRLI 2836

Qy 1321 TSPQOLASSKPMNPPSAANPCSPVQFSTPLAGLAKRRAGDPCGEMQSP-----TGLGQP 1377

Db 2837 TGTDPDLIVS--VTVPFSGHNRQNTASSVPSNSTSPILPSESITTPPPPTTTTAKP 2894

Qy 1378 K-----RRG-----RPPSKFFKQMEQRYQLTQAQVPPEMCSG 1411

Db 2895 ATTSCRGPPPSIQPAEMF-----TTTAPPPPSNG 2924

RESULT 19

QSHUMT

Microtubule-associated protein 2, splice form MAP-2b - human

N:Alternate names: MAP2

C:Species: Homo sapiens (man)

C>Date: 31-Mar-1989 #sequence revision 12-Apr-1996 #text_change 09-Jul-2004

C:Accession: I53693; A61085; F10024; S34131

R:Albala, J.S.; Kalcheva, N.; Shafit-Zagardo, B.

Gene 136, 377-378, 1993

A:Title: Characterization of the transcripts encoding two isoforms of human microtubule-associated protein 2

A:Reference number: I53693; MUID:94124038; PMID:8294038

A:Status: translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1824 <ALB>

A:Cross-references: UNIPROT:P11137; EMBL:7211958; GB:I12563; NID:gl850616; PIDN:AAB48098.

R:Dammerman, M.; Yen, S.H.; Shafit-Zagardo, B.

J. Neurosci. Res. 24, 487-495, 1989

A:Title: Sequence of a human MAP-2 region sharing epitopes with Alzheimer neurofibrillary tangles

A:Reference number: A61085; MUID:90096190; PMID:2481044

A:Accession: A61085

A:Status: not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 77-645 <DAM>

R:Kosik, K.S.; Orecchio, L.D.; Bakalis, S.; Duffy, L.; Neve, R.L.

J. Neurochem. 51, 587-598, 1988

A:Title: Partial sequence of MAP2 in the region of a shared epitope with Alzheimer neurofibrillary tangles

A:Reference number: F10024; MUID:88274407; PMID:2455776

A:Accession: F10024

A:Molecule type: mRNA

A:Residues: 489-1558 <KOS>

A:Cross-references: GB:M25668; NID:gl87380; PIDN:AAA59552.1; PID:gl87381

C:Comment: Microtubule-associated proteins are a complex group consisting of the high molecular weight proteins that stabilize the microtubules against depolymerization.

C:Genetics:

A:Gene: GDB:MAP2

A:Cross-references: GDB:118836; OMIM:157130

A:Map position: 2q34-q35

C:Superfamily: microtubule-associated protein MAP2b; MAP2/tau repeat homology

C:Keywords: alternative splicing; microtubule binding; phosphoprotein; tandem repeat

F:1455-1463/Region: microtubule binding #status predicted

F:1666-1696/Domain: MAP2/tau repeat homology <MT1>

F:1697-1727/Domain: MAP2/tau repeat homology <MT2>

F:1728-1759/Domain: MAP2/tau repeat homology <MT3>

F:1657,959,1064,1250,1436,1503/Binding site: phosphate (Thr) (covalent) #status predicted

F:1697,817,829,1320,1417,1542,1551/Binding site: phosphate (Ser) (covalent) #status predicted

Query Match 2.5%; Score 248; DB 1; Length 1824;

Best Local Similarity 19.8%; Pred. No. 0.0022;

Matches 317; Conservative 208; Mismatches 616; Indels 458; Gaps 75;

Qy 172 DSFPQNEEVCSGHPDEAKEMT-----SVVAE-----NGTGLVCSL 211

Db 493 DSFPVSLRQAVTDSAMTSKLTAKMTEPSALIEKSIQELFEMRVDDKDKIEGVGAATSA 552

Qy 212 ELEEQPELKMCGYNSVPSVSLHQSFLVLPDPVTSCLDDPSHLPLQLEDTPILSDDS 271

Db 553 ELDMPFYEDK-----SGMSKYFETSALKKEATSI--EPGSDYYELSDT---RESV 598

Qy 272 LEPNSLAPVPVSGGLYCIDDTTELMGAEDKULPLXDSVPVISALDCPSLNNATAFSLADDS 331

Db 599 HESIDTMSPMKNG-----DKFQTQKESQP-----SPPAQAGAGYSTLAQSPSDLPBPFS 649

Qy 332 QTSISIPASPTSPVVLGE-----SVLDNHSFDLNNSGDARQEBMET----- 372

Db 650 SPQERMTF--IDPKVYGEKRLHKNKDDLTLSRLSLGLGGRSAIEQRSMNSLNPMSCLDS 707

Qy 373 -----QSSDPPP-----SLTQAPADQSSTIQHLPATSPAVS-----PTTSPAVSLVVS 415

Db 708 IALGFNFGRGHDLSPLASDILTNTSGMDEGDDVLPATTPALEKAPCFPVESKEEQIEK 767

Qy 416 PAASPEISPEV---CPAA-----STVSPAVSVSVSPASS-AVLPVSLVPLTAS 462

Db 768 VKATGESSTQAEISCESPFLAKDFYKNGTVNAPOLPEMLDLAGTRSLASVSAEAVARR 827

Qy 463 VTSKASPVTSPPAAAPPTASPAKDV---SSFL-----TTADVEEITG 503

Db 828 KSVSETVEDSRGTGLPPVTDENHVIKTDLSQLEDLGYCVFNKYTVPLSPVQDSSENJSG 887

Qy 504 E-GLTASGSGDMRRRIATP---BEVRLPLQHGWRREVRIKKGSHRMQGETWYYPGCGKR 559

Db 888 ESGTFYEGTDKVRDLATDLSLIEVLAAGRKVDKDFSVK-----EASAHISGDKSGL 942

Qy 560 MKQPEVIK-----YLSRNLVSVRREHFSFRMPVGVDFEERDTPPEGLQWVQLSAB 612

Db 943 SKEFDQEKANDRLDTVLEKSEEHADSKHAKTTEE--AGDEIETFGI--GVTVRQALAK 998

Qy 613 E--TPSRQAITKRGPRNTEKATKEVPKVKSG-----RGRPPKVKITEL-----LN 659

Db 999 DLSIPTDASSEKAEKGLSSVPEIAEVEPSKVKVOGLDPAVQGO-LDVKISDFGOMASGLN 1057

Qy 660 KTDNRPLK-KLEAETETNEEDKAK-----IAKSKKKMKQKQVQSGECLTTIQQAENK 710

Db 1058 IDRRATELKLEATQDMTPSSKAPQEAADAFNGVESHGMEKTKVSETEV-----K 1107

Qy 711 RKQETSLKHKAEKKKXAEKKGTK---QEKLK-EKVREKKEKVKQKKEEVTAKK- 765

Db 1108 QKVAKPDVHQEAVDKESVSESGESHESLTMESLKADGKETSPESSLIQDEIAVKLSV 1167

Qy 766 ---PACKADKTIATQRLEERQOQMLEEMKPTEDMCLTDHQLPDPFSRVCGLTIPS 821

Db 1168 EIPCPAVSEADLAT---DERADVQM--EFIQGPKE-----ESKETPDISITP----- 1210

Qy 822 GAFSDCLTIVEFLHSFGKVLGFDPAKOVPSLGVQELGCLCGDSLGEVQDILLVRLKAL 881

Db 1211 ---SD---VAEPLH---ETIVSEPAE---IQSEEEIEAQ---GEYDKLLFR----- 1247

Qy 882 HDGFPSPYCSQSLKILGEKVSIEPLTRDNVSBILRCFLMAYGVXPALCDRLRTQPPQAQPP 941

Db 1248 -----SDTLQ--ITDLGVSGA-----REEFVETCP 1271

Qy 942 QOKAAVLAPVHELNGSTLIINEIDKLTLESMSYRKKNKWIVEGRLLRLTKVLAKTGRSE 1001

Db 1272 EHKGVIESVWTIEDFTTIVVQTTD-----EG-----ESSGHS 1304

Qy 1002 VEMGRPECLGRSSSRIMEETSGMBEESIAAVPGRGRRDGEDVDATASSIPELER 1061

Db 1305 VRPAALQPEVERPPSPHDEEFEEVEAAEAQ---AEP-----KDGSEAPAS--PEREE 1354

Qy 1062 QIEKLSKRQLF--FRKLLHSSQMLRAVSLGQDRYRRRYVLYPLAGIFVEGTEGNLVE 1119

Db 1355 VALSEYKTETVDDYDKDETTIDDSIMDADSL-----WV-----DTQDDRSIMTE 1398

Qy 1120 ---EVIKKETDSLKVAHAASLNPALPSMKMELAGNTTASSPARASRLTKTPGPMQPRH 1177

Db 1399 QLETIPKEEKAKEARRSSLE-----KHKKEKFPKTRGRISTP 1438

Qy 1178 FK-----SPVRGQDSEQPAQOL--OPEAQLHVPAOP---OPOLOLOSHKGFL 1221

Db 1439 RKVAKGPSTVSRDEVRKKGAVYKAEALAKKTEVQAHSPSRKFILKPAIKTRPHTLSCV 1498


```
Db 1794 TERHSTLSSAKTERHPVPSSKTEKHSVPSPSAKTERHSPASS--SKTEKHSVPSPS 1851
Qy 463 VTPSPKASPVTS--PAAAFPTASPAKNDVSSFLETTADVVEITGEGTLAGSGDVMRRRIAT 521
Db 1852 TKTERHSPVSSKTERHPVPSPSGK-----TDKRPVPSPSGKTEKHPVPS 1896
Qy 522 P--EEVRLPLQHGWRREVRIKKGSHRWOGETWY--GPCGKRMKQPFVIVIKYLSRLNVH 576
Db 1897 FORTKRLUPVSPSGRTD---KHQPVSTAGTEKHLPVSPSGKTEKQPP-----1941
Qy 577 SVRREHFSRPMVPGDFEERDTPGELQWQLSAEIIPRSIRIQAITGKRGRPRNTEKAKT 636
Db 1942 -----VSPTSKT-ERIEZ-----TMSVRELKMAFQS-----GDQPSKHKTGL 1977
Qy 637 KEVPKVKRGRGRPPKVIKTELLNKTDNRPLKKLEAQETLNEEDKAKIAKSKKGRKQKVR 696
Db 1978 FEHKSQKQ---KQPKQKGVKVEKEGPILITQREAAKTEN-----QTIKR 2019
Qy 697 GECLTTIQQOARNKRRKQETSKLKHKEAKKXAEKGTQKQELKEKVKREKKEKVKQK 756
Db 2020 GQRL--PVTGTAEKRGVRSIS---GVKCEDAAGKEKVLSHKIPFVQSPVEE--ESH 2072
Qy 757 EKEEYTKAKPACKADKTATQRLLEERQKQOMILEMKKPTEDMCLTDHQPDPDFSRYPG 816
Db 2073 RESEVPKEMKMA-----DEQGDMDLQISPRDKTSTDFSEVITKQELDNDRKYQQ 2119
Qy 817 LTLPSGAFSDCLITIVFELHS--FGKVLGFDPAKD--VPSLGLVQLGELLCQGSLSAGEVDLL 873
Db 2120 FRLSETEKAQLHLDQVLTSFNTTFFLDYMKDEFLPALS--LQSGAL--DGSS-----2169
Qy 874 VRLKAAHLHDGFPSPYQOSLKILGKVKSEIPLTRDNVSEILRCFLMAYGVXPALCDRLRT 933
Db 2170 -----ESLK--NEGVAGSP-----CGSLMEGTPTQISSEESYK 2199
Qy 934 QPFOAQPPQKAAVLAFPPVHELNGSTLIINEIDKT-----LESMSYKKNKIWVEGLRR 988
Db 2200 HEGLAETPETSPELSUSFPKKEEQTGETKSTKTETTEIRSEKHEPTTKDITGGSER 2259
Qy 989 LKTV-----LAKRTGRSEVEMGRP-----EECLGRRRSRIMEETSGMEEB---BEE 1032
Db 2260 GATVTEDESETSTESFQKEATLGSFKDTSRKQDDCTGSCSVALAKETPTGLTEEAACDEG 2319
Qy 1033 ESIAAVPGRGRGRDEEV--DATASS-----IPELEROIEKLSKRLQFRKKLHSSQML 1084
Db 2320 QRTFGSSAHKTQTDSEAEQESTATSDTKALPLPEASVKTDGTGTESK---PGCVIRSPQGL 2376
Qy 1085 RAVSLGQDRYRRYRWLVPLAGIFVEGTGKNLVPEVIKK--ETDSLKYAAHASL--NPAL 1141
Db 2377 ELALPSRD-----SEVL SAVADDSLAVSHKDSLEASVPL 2410
Qy 1142 FSKMELAGSNTTASSPARASRPPLKTRKPFMRHFKSPVRGQDSQFQAQLOPEAQL- 1200
Db 2411 -----EDNSHKTPDSLEPSPLK-----ESPCRDSLESSP---VEPKMAG 2448
Qy 1201 ----HVP---AQOPQLOLQIQLSHKGFL--BOEGSPLSGSQHDLSQSAFLSMLSQTQS 1251
Db 2449 IPPSHFPLPAAVAKTELLTEVASVRSLRLRDPDGSAAEDSLSEQTSLMESSGKSPLSPDTP 2508
Qy 1252 HSSLLSSSLVPLDSSPGKLDNAPSOPPEPEPEDEAES-----SPDLOAFWNTISAQM-- 1303
Db 2509 SSEEVSYEY--TPKTTDVS--TPKPAVTHECAEDDSENGEKKRFTPEEEMFKMVTIKRMD 2566
Qy 1304 -----PCNAAPTPLPPLAVSBDQF--TPSPQOLASSKPMNRPSAAN 1340
Db 2567 ELEQAKOKRDYKKEPKQBESSSSSDPDADCSVDVDEPKHTGSGED--ESGVPLVLTSESR 2625
Qy 1341 PCSPVQFSSTPLAGLAPKRRAGDPGEMPQSPGTGLGQPKRRGRPPSPKFFKQMEQRYLTQLT 1400
Db 2626 KVSSSSSESEPELAQL---KKGADSGLLP-----EPVIRVQPPS-----2660
Qy 1401 AQPVPEPCSGWWIIPDDEMIDMLKALHPRGIREKALHKHLNKHRDFOEVCULRPPSADP 1460
Db 2661 --PLPSSMDSN-----SSPE-----EVQFPVVSKQYTF-----KWNEDT 2693
```

```
Qy 1461 IFBFRQLPAPFOEGINSMSPKPKTYETDLAVLQWVEELEQRVIMSDLQIRGWTCPSPDSTR 1520
Db 2694 QEEPCK-----SEEEKDSHSLA-----EDRHAVS-----TEA 2721
Qy 1521 EDLAYCEHLSQSDIEDITWRGPGRECLAPQRTTNPLDLAVMLRLAALAEQNVKRYLREPLW 1580
Db 2722 EDRSYDKLNRDTPDKICDGHGCEAMSPS--SARPVSSGLQS-----2762
Qy 1581 PTHEVLEK-----ALLSTPNGAPEGTTEISYEITPRIRIWRQTLQRCRSHAHVCL 1632
Db 2763 PTGDDVDQEPVIYKESLALQGTHEKDTGEELDVSRAESPOADCFSESSSSSLPHCLIV 2822
Qy 1633 CLG-HLERSIAWEKSVNK 1649
Db 2823 SEGKELDEDISATSSIQK 2840

RESULT 21
TS1023
hypoetical protein B7F21.40 [imported] - Neurospora crassa
C:Species: Neurospora crassa
C>Date: 21-Jul-2000 #sequence_revision 21-Jul-2000 #text_change 09-Jul-2004
C:Accession: TS1023
R:Schulze, U.; Aign, V.; Hoheisel, J.; Brandt, P.; Partmann, B.; Holland, R.; Nyakatura,
submitted to the Protein Sequence Database, July 2000
A:Reference number: 225286
A:Accession: TS1023
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2649 <SCH>
A:Cross-references: UNIPROT.Q93J0; EMBL:AL389901; GSPDB:GN00116; NCSP:B7F21.40
A:Experimental source: BAC clone B7F21; strain OR74A
C:Genetics:
A:Gene: NCSP:B7F21.40
A:Map position: 6
A:Introns: 1619/3; 2584/1

Query Match 2.5%; Score 245.5; DB 2; Length 2649;
Best Local Similarity 18.1%; Pred. No. 0.0046;
Matches 318; Conservative 182; Mismatches 650; Indels 609; Gaps 68;

Qy 97 NYSQPSANPSNL---KDPPL-----LSQFSGGQVPLNGILGGRSQSPSPSHNTNLRAGS 149
Db 503 NIERGPTGPSKPLLPESGPPIPVGPRAQRSSKQW-INPTL-GKKIPESPK-----SARS 555
Qy 150 QKFWANGTHSPWGLNFDQSQELYDSFPQNFEEVCSGIHPDEAAEKEMTSVVAENGTGLVC 209
Db 556 QSFVSQSOH---GRPFGHRPESHSDHHGETERPRSPDAKSEHLVAADQGSRSLLT 612
Qy 210 SLELEEXQPE-----LXMCYNGSVSPVESLHQE-----VSVL 242
Db 613 GANDTPRPERGTOSARASVDRTFRASLDNDRMTDSFGGAHRSFRELDAERDRERPAIM 672
Qy 243 -----VPDPTV-----SCLDDPS-----HLPDQLEDTPILSEDSLEPFNSLA 279
Db 673 APMQREDADKSRIPESRVPVKRKRSLVAIPSTRLQLPTQTSLPGTESDDDEDMDDYF 732
Qy 280 PEPVSGGLYDIDTELKGAEDKLPXDSFVLSALDCP-----SLNATAFSLA--- 328
Db 733 DAEITK-----QETELKKLDKSTAGVPMQIIVRQVACTVHDAMVKVSVSLVDMIGGLP 787
Qy 329 -----DDSTSTSIASPTSPPVLGESVLDNNGSDABQ 367
Db 788 EGYSPRPVPAKHTQKESPOPKQEQELQFSAKPKRPV-----QEAVPAVAQOPTEQ 841
Qy 368 EEMETQ-----SSDPPPSLTQAPQDSSTIQLHRTSPSPAVSTTSPA 409
Db 842 VEKPVQKVPQKPAOKLDELEPPVSESEPEVAPKEPEPKPVSEPKKAPPELLPEKLP 901
Qy 410 VSLVSVPAASFEISPEVCPAALSTVSPAVSVSPASSAVLPAVSLVPLTASVTSFKAS 469
Db 902 FEQIPRPQFPQFPQFQQAQAPQQLLTAEPKSVTEPEREREAEVEQEAETQBOA 961
```

```
Qy 470 PVTSPAAAFPTAS----- 482
Db 962 PTERQLLAIPTEVNEETSFLRPPQPKEEMDTGGGSLPLPLTVVEHKVVDQVDMVDV 1021
Qy 483 -----PANK-----DVSSFLETTADVEEITGELT 507
Db 1022 DDEGRDALQPPCKVPPLRDVSDGESVSQFPRPNQPTSDGASSRMDDESEERTEDDAS 1081
Qy 508 ASGGSDVMRRRIATPEEVRLLPQH--GWRREVR1KKGSHRWOG--ETWYWG-----PCG 557
Db 1082 VSGSVHVRRELSSTPTDELFPVFNKPHQSKVKRLSEQSPGFGAFPMGNIIRDMAETA 1141
Qy 558 KRMQKQPEVIK-----YL-----SPNLV-----HSVRRHFSP 587
Db 1142 KEQHEAKQVYRTNVDAYLRLFTLSDPTAVKSRNQFSQSDKDKVSGKGHSGDNHKEGG 1201
Qy 588 RMPVGDFFPEERDTPGELQVOLSABEIPSR1QAITGKRGRPRNTEKAKTEKVPKVGKRG 647
Db 1202 RRTTSRSTELDVVAIQESIREAQEKERE--RAQREKYRTDKEAVIPEMIWTDEKD 1259
Qy 648 RPPKVKITELNKTNRPLKL-----EAQETLNEDKAKIA-----KSKKMKRQK 693
Db 1260 HQLFVDTSGLL-----PLEKLVTQAVPYHVNFATBAEAEKFEKAVLEYKPKQWKIAHE 1313
Qy 694 VORGECLTTIQOARNKQETKSLKHKEAKKXAEKEGKTQ-----EKL 741
Db 1314 LPNDRDFHSVIQYYAKKRE-----LNLKERLKKQPRRRKGRGKQKNALVSELGNPENE 1368
Qy 742 KEVKREKEKVKMKEBEVTKAKPAC-----KADKTLAT-----QRRL----- 780
Db 1369 TTEDNQENGNSGRQOPRAAAPSNGHEATPNADSDGATPSATPGRRRAGITTEPKND 1428
Qy 781 -----BERQKQOMILEEMKPTEDMCITDHPDPDFSRVPLGLTLPSGAFSACL 828
Db 1429 SGAEKPEGVKGGRRARQKADKEPKVPKPAQAITA--PTPPAASGKGTNARSNSRNV 1485
Qy 829 TIVEFLHSFGVLGFDPAKDVPSLGVLOEGLLCOGDSLGEVODLLVRLKLAALHDPGPPS 888
Db 1486 QNPEWM-----PKTPAEGL-----ARVONPMDFVPPGSMQFPLAPA 1522
Qy 889 YCQSLKILGEKVSBEIPTRDNVSBILRFLMAYGVXPALCDRLRTQPPQAOQPQOKAAVL 948
Db 1523 --QOTPLASPERAPTLASTTISEVM-----APPSL-----RPEFPAPPASLP 1563
Qy 949 APVHEUNGSTLIINEIDKTLIESVSSYKNKWIVE-----GRLLRLKT----- 991
Db 1564 TFEIGQSSGPRI-----RTPQQASSY-----WSVSESNDPFLSLRSFGTDWSA1ANHMT 1614
Qy 992 -----VLAKRTGSEVEMGRPECLGRRSSR-----IMEETSG- 1025
Db 1615 KQVMVRNYLQRKKEGQEQE1ATEADLKQKQGERRRPPPTPSAGPRKRYDVPSSGH 1674
Qy 1026 -----MEEBEEBES1AAVPGR----- 1042
Db 1675 RPLAAAEAEQPAKSEAPANQFSPFQVTP1QASPVLT1VQRPSTVMPAPLPAPV 1734
Qy 1043 -----GRRDGEVDATASS1PELBERQIEKLSKRLQFPRKLLHSSQMLRAVSL 1089
Db 1735 AASASAAPSQTPSRKPGAVSQTMSVPBHLRQ----- 1767
Qy 1090 QODRYRRRYW1PVLAG1FVETGTEGNLVPBEV1KETD-----SLKVA-AHAS1N1PALP 1142
Db 1768 -----PTATFVFTE-REGEP1PVSQPAQTQHQPVRI1SOKTAPVPSMSFAS 1814
Qy 1143 SMK1ELAGSNTTASSPA-----BARSRLKTKPGFQMPRHFKSPVRQD 1186
Db 1815 EAMPSAGW1FEANRP1TLLSQQHELREVR1REGRE1QIN1EP--1QPR1QPR1PM----- 1868
Qy 1187 SEQPQAQL--OPEAQL-HVPAQ--PQQLQ1QLQSHKGFLBQEGSP1LSGOSHD1LSQA 1241
Db 1869 -ERPAMRFKQEP1QHLHDP1QNFMP1QORS1QMPRA-----EAA1PMGRQD1PPRS1MAPSA 1920
```

```
Qy 1242 FLSWLS--QTQSHSSLLSSSVLTPDSSPGKLDPAPOPP-----EPEPEDEAESPDLOAF 1295
Db 1221 POSTTPIQAPVYNLLSESV-PPORTPQPSPAWERPPVPSTORPMTSNQ-----EQY 1974
Qy 1296 WFNISAQMP-----NAAPTPPLAVS-----EDQTPSPQOLASSKPMNRPSAANP 1341
Db 1975 SANTSAAPVPPPPQAPPAIPPEPKKVSSIFSILNDDNPAPAPAPKRVNDVASMPRAAS- 2033
Qy 1342 CSPVQFSSTPLAGLAPKRRAGDPCQEMQSPSTGLCQPKRG-----RPPSKFFKQM 1391
Db 2034 -----TSTP-----PQOMSARPPQPPPTTAVSASQRDRTEALGYSYARNPPPSAAQAAM 2083
Qy 1392 E--ORYLTOLTAQVPPPEM 1408
Db 2084 PSLAPYHTQ---SPQPHM 2099

RESULT 22
T29757
Protein UNC-89 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 03-Dec-1999
C:Accession: T29757
R:Du, Z.; Le, T.T.; Wilson, R.
submitted to the EMBL Data Library, May 1997
A:Description: The sequence of C. elegans cosmid C09D1.
A:Reference number: Z20679
A:Accession: T29757
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-6642 <DUZ>
A:Cross-references: EMBL:AF003131; PIDN:AAB54132.1; GSPDB:GN00019; CESP:unc-89
A:Experimental source: strain Bristol N2; clone C09D1
C:Genetics:
A:Gene: CESP:unc-89
A:Map position: 1
A:Introns: 17/2; 108/3; 154/2; 211/2; 265/3; 326/2; 352/3; 426/2; 454/1; 500/1; 537/1; 537/1; 6027/1; 6061/3; 6153/2; 6515/1; 6552/3; 6609/1

Query Match 2.5%; Score 244.5; DB 2; Length 6642;
Best Local Similarity 18.7%; Pred. No. 0.016;
Matches 338; Conservative 230; Mismatches 685; Indels 559; Gaps 77;

Qy 198 SVVAENGTGLVCSLELEEXQPELXKCGYNGSVSPVESLHQEVSVLPDPTVSCLDLDPHSL 257
Db 1219 SWAEN-----TLGAEETGAQL-----TIEP--KKESVVVEKQDLSSEVQKEI 1260
Qy 258 PQDLEDTPILSEDSLEFPNSLAPFVPSGGLYGIDDTLGMGAEDKLPLKXDSV1SALDCPS 317
Db 1261 AQQKEA-----SPEA-----TTTITMETS1TSTKTTMTTETS 1296
Qy 318 LNNATFSL1ADDSQTS1FASPTSPVLGE-SVLQ-----DNSFDLNG-----S 363
Db 1297 TVGGVTVEKSESESAT1VIGGSGGVTEG1SVSKIEVVSKTDSQTDVREGTPKRRVS 1356
Qy 364 DAEGE-EMETQSSDFPSLTQPADQDST1QLHPATSPAVSP1TSPAVSLVVS1PAASPEI 422
Db 1357 FAEELPKVEIDSD--RKKKSPSPDKKEKSPKEETKEPA--SPTKKTGEE--VKSPEKSP 1413
Qy 423 SP---EVC1PAASTVVS1PAV7SVSVSPASSAVLPAVS1LEVPLT-----ASVTSPKASPV 471
Db 1414 SPTKKEKSPAAE1EVKSPTK--KEKSPSS-----PTKKEKSPSP1TKTGDEVKEKSPKSP 1468
Qy 472 ---TSPAAAFPTAS1PANK-----DVSSFLETTADVEE--1TGEGLTASGSDVNR 516
Db 1469 KKEKSPKEP1EDVKSPVKKESPDATN1VEVSS--ETTIEKTETMTMTMTMTES1ESRTSV 1526
Qy 517 RRIATPEV-----RLPL-----OHGWRREVR--1KKGSHRWOGETWY1GCGK--RMKQF 563
Db 1527 KKEKTEPKVE1DKSPKTPKDKSPKSPK1TEBKSPVKKKESPEKVEEKSPKSPK1KESPEK 1586
Qy 564 PEV1KYL1RNLVHS1VRREH1FSFS1RMP1VGDF1FERD1TPEGLQW1QLSAE1IPSR1QAITG 623
```



```
Db 2156 -NPTYPEG-----QPSGNLAQYCPAASQATAVRQLLPSTATVRAADGMIYSTINTPI 2207
Qy 347 LGESVLQDNFNLNNGSDABQSEMETQSSDFPPLSTQP-----APQSSSTIQLHPATS 399
Db 2208 -----AATLPIITQ-----PASVLRPMVRGMYRYPVYSGGVTAAPLTS 2245
Qy 400 PAVSPPTSPAVSLVVS-----PAA SPEISPEVCPAASTVV--SPA VFSVVS PASSA----- 448
Db 2246 LTRVPMIAPRVPGLPAGLYRYPAPRPFIASSVPPAEGPYLGPAP-----ATKASGAGGPP 2301
Qy 449 --VLPA-VSLEVPPLTASVTSPP--KASPVTS PAAAFPTASPANKDVVSFLETTADVEBIT 502
Db 2302 RPPLPAGVAREEPP--SITAPAVIKEAPVAPAGPAPAPPPQCKP-----A 2345
Qy 503 GEGLTASGSDVMRRRIATPEEV-----RLPLQHGWR 535
Db 2346 GEAAAGSGG-VLSRPA SEKEASQEDRQRKQEQQLQLERERVELEKRLQLRLOEELER 2404
Qy 536 EVRIKKGSHRWGETWYWGCGKRMKQPPVIVKYLRSNLVHSVRREHFSFSPRMPVGDFF 595
Db 2405 E-RVSLQRHREBEQLLV-----QRELOELQTIKQ-----HVLOQQQ----- 2439
Qy 596 EERDPPEGLQWVOLS-----AEIEPSRIQALTKGRGRPRNTEKA---KTKEVKPKYKRG 647
Db 2440 EERQAPALOREQLAQORLQLEQIQLOLQOLQLEBEQKQKAPFPATCEAP-----SRG 2495
Qy 648 RPPKVKITELNKT-----NRPLKLEAQETLNEDK 680
Db 2496 PPPAA--TELAQNGYWPPLTHAAFI AVAGTEGCGQPREPVLHRLGLPSSADMSLOTBEQ 2553
Qy 681 AKIAKSKKMKQKQVOR-----GEC LTTIQOQARNKQKQETKS-----LKHKEAK 724
Db 2554 WEAGRSIGIKGHSMPRLDACEPEGSDPSTVRR IADSSVQTDDEGEGRYLVTRRRNR 2613
Qy 725 KKSXAEKKGKTKQEKLEKVKREKKEVKMEKEEVTAKPACADKTLATORLEERQ 784
Db 2614 RSADCSVQTDDEADNADWQPVRRRRSRLSRHS DSGSDKDATASSSTTAATAKAMSSV 2673
Qy 785 KQOMLEEMKPTEDMCMLTDHQP LDFSRVGLTLPLSGAFSDCLTIVEFLHSFGKVLGFD 844
Db 2674 GIQTI-----SDCSVQTEPEQLPRVS--PAIHI-TAATDPKVEIYVI----- 2713
Qy 845 PAKOVPSLGVLOEGLLQCGDSLGEVQDILL-VRLKKAALHDPFSPYCSOLKI-----LG- 897
Db 2714 ---SAPEKTGRGESLACQTEPDGOAQGVAGPQLIGPTAISP-----YLPGIQIVTPGALGR 2766
Qy 898 -EKVSEIPLTRDNVSEILRCFLMAYGVYPALCDRLRTQPFQAQPOQKAAVLAPVHVLN 956
Db 2767 FEKKKPDPL-----EI-----GYOAHLPPELSQLVSRQPP-KSPQVLYSPVSPLS 2811
Qy 957 GSTLIINEIDKLTESMSYRNKNKIVEGRLRLKTVLAKRTGRSVEVMGRPEECIGRRRS 1016
Db 2812 PHRLDTSF-----ASSERLNKAHVSO---KQFIADSTLRQOT-LPRMKT LQRLS 2860
Qy 1017 SRIMEETSGMBEEBEEESIAAVPGRGRRR-----KQFIADSTLRQOT-LPRMKT LQRLS 2860
Db 2861 DPKPLSPTEASSAKERFSLYHQGGLGQSVLPNPNGLVRKVRTLPSPPPPEEAHLPLAG 2920
Qy 1048 EVDA-----TASSIPELERQIE-----KLSKRO----- 1070
Db 2921 QVPSQLYAASLLORGLAGTTPVATKASLLRELDRLRVEHSTKLKAKQALDEEKE 2980
Qy 1071 LFFRKLLHSSQMLRAVSLGQDVRVRRVYVLPYLAGI-----FVEGTE-----GNLVP 1118
Db 2981 IDAKLYLELIGITORKESLAKDRGRDY---PPLRGLGEHRDYLSDSELNQLRLOGCTTP 3037
Qy 1119 BEVIKKTDSLKVAHAHSLNPA LFSMKM--ELAGSNNTTASSPARARSPLTKPKGFMPOR 1176
Db 3038 AGQVVDYVPA SAAVPATPS-GPTAQOQPFPPAAQYTAGSSGPTQNGFPFAHQAPTYTCPS 3096
Qy 1177 HFKSPVRQDSEQP-----QAOLQPEALQHVPAQPOQLQ-----LQLOSHKGFLQSGS 1226
Db 3097 TYPARTYPPGTGYPAEPLGSLPQAPFHPGTGHVAAPT-PMPTTQSAFFPVQADSRAAHQKPR 3155
```

```
Qy 1227 PLSLQSQSHDLSQS AFLSMLSQTQSHSLSSSVLTTPDSSP---GKLDPA PS-----QP 1277
Db 3156 QTSLA-----DLEQ-----KVP TNYEIVGSPAVTMSSAPPETGYSGFAVSGSYEQGA 3203
Qy 1278 PEEPEPDABESSPDLOAFWFNISAQPCNAAPTTPPLAVSEDOPTPSPQOLASSKPMNRPS 1337
Db 3204 PEHPRGSDRSS-----VSQSP---APTYP---SDSHYTSLEQNVPRNYM----- 3242
Qy 1338 AANPCSPVQFSGSTPLAG-----LAPKRRAGDPCMPQSPGTGLGQP----- 1377
Db 3243 -IDDISELTKUSTPTASESQRLPLGPGVSGRPGKDPGEPAVLGEGPTLPCCYGRGEBBS 3301
Qy 1378 -----KRRGRPPSKFFKQMEQRY---LTQLTAQVPPPEMCSGWWIWD 1417
Db 3302 EEDSYDPRGKSGHRS MESNGRPS THYVGDSDYRHGARADKYGPGMPGPKHFS----- 3354
Qy 1418 PEMLDAMLKALHPRGIREKALHKLHKNRDFLQEVCLRPSADPIPEPQLPAFOGIMSW 1477
Db 3355 -----KSLAPAAISSKR-----SKRK-----QGMQEKISKF 3381
Qy 1478 SPKE--KTYETDLAVLQWVEELEQRVIMSDLIQIRG-----WTCP 1514
Db 3382 SPIEAKVDES DLA-----SYPPPTVSSSITSRGRKFQDEITYGLKKNVYEQQRYGVS 3435
Qy 1515 SPDSTRED-----LAYC-EHLSDSQEDITWRGPGREG LAPQRKTTNPLDLAV 1560
Db 3436 SRDAAEEDERMVSGSSRSR MASAYSGEKL--SSHDYSSRGKGYE---RERDTAERLQKAG 3490
Qy 1561 MRLAALQNVKRRYLREPLWPTHEVLEKALLTPNGAPEGTTTISYEITPRIRWQT 1620
Db 3491 SKPSSLSMAHGR--ARPPM-----RSQASSEESPVSP----- 3520
Qy 1621 LQRCRSAHVCLICLGHLEERSIAWEKSVNKTCLVCRKDNDEFULLCDGCRGCHIYCHR 1680
Db 3521 LGRPRPAG-----GALPPG-----DTCQFCSSHMP 3547
Qy 1681 PKMEAVPEG-----DWFCTVCLAQ-----QVEGEFTQKPGPKRQK-KRK 1719
Db 3548 DVQEHVKDGPRAHAYKREEGYMLDSDSHCVSDSEAYHLGQEBTDWFDKPDARSDFRHH 3607
Qy 1720 SGYSLNFSEGGRRRRVLLKGESPAAGPRYSEERLSPSKRRRLSMRHHSD 1771
Db 3608 GGHTVSSSQKGRPARHSYHYDEPPEEG-LWPHDEGGGGRHTSAKEHRHSD 3658

RESULT 25
T00363
hypothetical protein KIAA0674 - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 09-Jul-2004
C:Accession: T00363
R: Ishikawa, K.; Nagase, T.; Suyama, M.; Miyajima, N.; Tanaka, A.; Kotani, H.; Nomura, N.
DNA Res. 5, 169-176, 1998
A:Title: Prediction of the coding sequences of unidentified human genes. X. The complete
A:Reference number: 214142; MUID:98403880; PMID:9734811
A:Accession: T00363
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1234 <ISH>
A:Cross-references: UNIPROT:Q9Y4D0; EMBL:AB014574; NID:g3327161; PIDN:BAA31649.1; PID:g3
A:Experimental source: brain; clone HK02519
C:Genetics:
A:Note: KIAA0674
F:212-263/Domain: BKBP-type peptidylprolyl isomerase homology <PRI>

Query Match 2.5%; Score 241.5; DB 2; Length 1234;
Beet Local Similarity 19.6%; Pred. No. 0.0026;
Matches 238; Conservative 169; Mismatches 415; Indels 395; Gaps 56;

Qy 247 TVSCLDL--PSHLP--DQLEDTPLSDESLPFFNSLAPEPVSGGLYGDIDTLMGAEDKL 302
Db 316 SVSSRDSAAAPSPIGADNLSADPVVSPPTSIPFKS--GEPAL-----RTKSNLSLSQL 366
```



```
QY 846 AKDVPSLVGLOGLLCOGDSIGEVDLRLVLLKAAALHDPGF-----PSYCQSLK 894
Db 500 ASAAPMOG-----KKSILREPTFRWTSLKHSRSEPOVFSSAK 537
QY 895 ILGE-----KVSEIPLTRDNV-----SEI 913
Db 538 YAKEGLIKRPIDNRPPLTPEDVGFASGFSASGTAASARLFSPLHSGTFRDIHKRSPI 597
QY 914 LRC-----FLMAYGVXPALCDRLRTQPPQAO----- 939
Db 598 LRAPFTPEAHSRIFESVTLPSNRTSSGASSGVSNRKRKRVFSPIRSEPRSPSHMR 657
QY 940 -----PPOOKAAVLAFPVHELNGSTL-----IINIDKLTLESMSYRKNKI 981
Db 658 TRSGRLSTSELSPLTPPSSVSSLSIPVSLAASALNPTFTFPHSHLTQSGDSTEKNQ-- 715
QY 982 VEGRLRLKTLVAKR-----TGRSEVENGR-----PRECLGRRSSRIMETS 1024
Db 716 ---RARKOTSAPAEFSSNSPALFPWFTPGSGTEKGRKDDTAPELSKRDADKSVEKD 772
QY 1025 GMEEEEEESTAAVPRGR--DGEVDATASSIPELEROIEKL-----SKROLFF 1073
Db 773 SRERDREKENKRESRKEKKKSGDIOSSSALYPVGRVSKKVGEDVGTSSSAKATG 832
QY 1074 RKK--LLHSSQMLRAVSLGQDRYRRYVWLPLYLAGIFVEGTEGNL-----VPEEVI 1122
Db 833 RKKSSSLDGGADVAPVTLGD-----TTAVKAKILIKKGRNLEKNLNLGDLGPAAPSL 884
QY 1123 KKETDSLKVAHASLNPALFMSKMWELAGSNTTASSPARARPLTKPGFQWPRHFKS-- 1180
Db 885 KERTPCLSAPESSVTKHSTSSIGSLAQADKLPMTDKREVASLLKAKAQLCKIEKSLK 944
QY 1181 ----PVRGQDSEQOALOPEAQLHV-----PAQPOQLQLQSHKGF 1220
Db 945 QTDQPAQGSDDSETSVRGRIKHVYRRAAVALGRKRAVFPDMDPTLSALPWEEREKI 1004
QY 1221 LEQEGS-----P 1227
Db 1005 LSSMGNDKSSVAGSEDAEPLAPTIKPKVTRNKAPQEPVKKGRRRRRCGQCPGQVP 1064
QY 1228 LSLGQSHDLQSASF-----LSWL-----SQTQ 1250
Db 1065 EDCGICTNCLDKPKFGGRNKKQCKMRKQNLQWMPKSKALQKTKAVKKEKKSKTTE 1124
QY 1251 SHSLLSSSVLTPOSSPKCLDPAPSQP--EEPEDEASSPDLOAFWFENISAQMPGNA 1307
Db 1125 KKESEKSTSVKSP-----LEPAKAAAPPREEPAPKSSSEPPPRK--PVEKSEBGG 1175
QY 1308 APTPLAVSEDOPTSPQOLA-----SSKPMNRPSAANPCSPVQFSSTPLAGLAPKRA 1361
Db 1176 APAP-----APAPEPKVASPASRKSQVSPAAVVPQP-----PSTAPQKE 1220
QY 1362 GDEMPPOSPTGLQPKRGRRPPSKFKOMQORYLTQTLTAOPV-----PDMCSGW 1412
Db 1221 A-PAVAPSEP-----KKQPPPPPEPGPQSKQKVAPLPSIPVKQPKDKKEKPP----- 1268
QY 1413 WNIIDPEMLDA-MLKALHP--RGIREKALHKLNKR--DFLOEVCLRPSPADTIFEPHQ 1466
Db 1269 ---PVSKQENAGTNILNPLNGISSKOKIPADGVHRIKRVDP-KEDC---BAENNVWEMG 1321
QY 1467 LPAFOEGIMSPKEKTEYTDLAVLQWVEELEQRVIMSDLIQIRGWTCPSPDSTREDLAYC 1526
Db 1322 L-----GILTSVP-----ITPRVVC-----PLCSS--SEHVEFYVC 1350
QY 1527 E-----HLSSQEDITWRG-----PGREGIAPQRKTTNPLDLAVMRLA 1564
Db 1351 QVCCEPHFKFLEENERPLEQLENWCCRCKFCHVCGRQ-----HOATKQLLECNKCRNS 1406
QY 1565 ALEQNVKRYLREP-----LWPTHEVLEKALLSTFNGAPGTTTSEYITPRIRIWRQ 1619
Db 1407 YHPECLGNPYTKPTKKKWKVICTKCVCKRSCGSTTPG--KGWDAQMSHDF----- 1455
```

```
QY 1620 TLQRCRSAHVCLCL---GHLERSIAMKESVNVKVTCLVCRK-GONDEF---LLLCDGCDR 1672
Db 1456 -----SLCHDCAKLFAGNF-----CPLCDKCYDDDDYESMMQCGKCDR 1495
QY 1673 GCHYIYCHRPK-----MEAVPEG-DWFCTVC-----LAQQVEGFTQKPGPK 1713
Db 1496 WWHKSCSLSTEDBEMTEILNSLPESVAITCVNCTERHPPEWRLEALEKQLASLKQVITA 1555
QY 1714 RGQKRKSGYSLNFGSEGDGRRRRVLLKGRSPAAGPRYSE---ERLSPSKRRRLSMRNHHS 1770
Db 1556 LLNSRTTSHLLRYRQ-----AAKPDLPNTEESIPSRSSPE----- 1592
QY 1771 DLTCEIILMEMESHDAAPPEXPNPRLVSGYRIIKNPMDFMTMRRLRLRGVTSSEE 1830
Db 1593 -----GDPDPVLTVEVSKQDEQQPLDLEGVKRMDQGSYVSVLE 1630
QY 1831 FAADALLVDFNCQTFNEDDSEVGKAGHIMRPFESRWEFFY 1871
Db 1631 FSDDIVKIIQAIINSDGGQPEIKKANSMVKVSFFIRQMERVF 1671

RESULT 27
T42761
Bassoon protein - rat
N:Alternate names: brain-specific synapse-associated protein
C:Species: Rattus norvegicus (Norway rat)
C>Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 09-Jul-2004
C:Accession: T42761
R:Dieck, S.; Sanmarti-Vila, L.; Langnaese, K.; Richter, K.; Soyke, A.; Wex, J. Cell Biol. 142, 499-509, 1998
A:Title: Bassoon, a novel zinc-finger CAG/Glutamine-repeat protein selectively localized
A:Reference number: 222249; MUID:98345363; PMID:9679147
A:Accession: T42761
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-3938 <DIE>
A:Cross-references: UNIPROT:O88778; EMBL:Y16563; NID:G3413503; PIDN:CAA76287.1; PID:G341
A:Experimental source: strain Sprague Dawley; brain
C:Function:
A:Description: may be involved in cytomatrix organization at the site of neurotransmitter
A:Note: Component of the presynaptic cytoskeleton
C:Keywords: coiled coil; zinc finger

Query Match 2.4%; Score 240.5; DB 2; Length 3938;
Best Local Similarity 19.3%; Pred. No. 0.013;
Matches 352; Conservative 200; Mismatches 655; Indels 617; Gaps 86;

QY 16 GLPPAPAAAGLKPSPSS-----GEGLYTNGSPMNFPOQKSLNGDV 56
Db 1825 GPVPEPGAEPHRATPAELRAHALPGTRKPTVVVVQMGGAAGTVTTL-LPEEPA---GAL 1880
QY 57 NVNGLSTVSHTTTSGILNSAPHSSSTSHLHHPSVAYDCLWNYSQYPSANPGSNLKDPPLL 116
Db 1881 DLTGMRPESRLAACCDMAYKFPFGSSCTGTFHPA-----PSA-PKSVTDAALP 1927
QY 117 SQFSGGYPLNGILGSGRSQSSPSHNTNLRAGSKQFWANGTHS-----PMGLNFDQSLEY 171
Db 1928 QSSSGPYF-----SPRDPPEPPLTFRAQGVGPGPHEBQRPYPQGL---PGRLY 1974
QY 172 DSFPDQNFEEVCSGIHDEAAEKMTSVVAENGTLVCSLELEXQPELKM-----CG 224
Db 1975 SSMSDNTLAEAGLNYHAQRIGQ-----LFQGPGRDSAVDLSSLKHSYSLGFADGRYLG 2027
QY 225 YNGSVPSVESLHQEVSVL---VPDPTVSCDDD---PSHLPDQLEDTPTLSEDSLEPFNSLA 279
Db 2028 QGLQYGSFTDLRHPTDLSLHPLPMRPYSSVSNYSVDHRYGRPRGDAVGQEAASLAQYSAT 2087
QY 280 PEPVS-----GGLYGIDDTLMGDAEDKLPLKXDSPVISALDCPSLNNATAFSL 327
Db 2088 AREISRMCAALNSMDQYGGRRGGGS---GGPDLVFPYQPQ-----HCFGLNAPQGLASL 2137
QY 328 ADDSQTSTSIASPTSPVVLGESVLQDNISFDLNNGSDAEQBEMETQSSDFPPSLTQ--PA 385
```

```
Db 2138 -----RSLGLGNPTYP-----EQGPS--PGNLAQYCPA 2163
Qy 386 PQSSTI-OLHPATS-----PAVSPTTS----- 407
Db 2164 ASQGTAVQLPLSATVRAADGMISTINTPIATPLTTPASVLRPMVRGMYRPGS 2223
Qy 408 -----PAVSLVWSPAASPEI-----SPEVCPAASVTVPVAFSVV--SPASSA 448
Db 2224 GGVTAVPLTSLTRVPMIAPRVPLPAGLYRYPASRPPIAST-IPPAEGPVYLGKPAAK 2282
Qy 449 V-----LPA-VSLEVPMTASVSPKASVPTSPAAAPTASPANKOVSSFLETTAD 497
Db 2283 ASGAGGPPRPPLPAGGAREEPL--STTAPPAVKEAPVAQAPAPPQGP 2330
Qy 498 VEEITGELTAGSGDVMRRRIATPEEV-----RLPLQ 530
Db 2331 ----AGDAAGSGG-VLGRFVMEKEASQEDRQKQEQQLQLLERERVELEKURLQLQ 2385
Qy 531 HGWREVRIKGSHRWQGETWYGPCKRMKQFPPEVIKYLNRNLVHSVRREHFSFSPMP 590
Db 2386 EELERE-RVELQRHREEQLLV-----QRELEQLQTIKH-----HVLQOQQ----- 2425
Qy 591 VGDFFEEDTPEGLOWQLS-----AEEIPRIQAITGKRGRPRNTEKA-----KTKEVPKV 642
Db 2426 ----EERQAQALQREQLAQRLEQIQLOLQLOQLQLEEQKQKQKAPPAATCEAP-- 2478
Qy 643 KRGRGRPKVKITELLKNTD-----NRPLKLEBAQETL 675
Db 2479 --SRGPPPA--TELAQNGQYWPPLTHTAFTAVAGTEGPGQARBVPVLRHGLPSSADMSL 2534
Qy 676 NEEDKAKIAKSKWQKQVQVQ-----CECLTTIQGQARNKQKQETKS-----LK 719
Db 2535 QTEQWENAGRGIKKRIUMPLRLDACEPESGPDSTVVRRIADSSVQTDDEGEGRYLLTR 2594
Qy 720 HKEAKKSXAEBKGTQKEKLKVKREKKEKVEVTKAKPACKADKTLATQRR 779
Db 2595 RRRTRRSADCSVQTDDEDAEWEQPVRRRRSRLSRHSDSGSDSKHEASASSAAAAARA 2654
Qy 780 LEEKQKQMIIEBKKTEDMCLTDHQLPDPFSVPGLTLPSGAFSDCLTIVFLHSGK 839
Db 2655 MSSVGIQTI-----SDCSVQTEPEQLPRVS--PAIHI--TAATDPKVEIVRYI----- 2698
Qy 840 VLGFDPKADVPVLGVLOEGLLCCQDSLGEVQDLL-VRLKKAALHDPGPSPYCSQSLKI--- 895
Db 2699 -----SABEKTGRGESLACQTEPDGQAGVAGPQLIGTAISP-----YLPQIQTVP 2746
Qy 896 --LG--EKVSEIPITRDNVSHILCFLMAYGVXPALCDRLRTQFPQAPPOCKAAVLAFP 951
Db 2747 GALTGRFEKKPDPL-----EI-----GYAHLPEPSLSQLVSRQP--KSPQVLYSP 2791
Qy 952 VHELNGSTLIINEIDKTLSESSYRKQKWIVEGRRLRLKTVLAKRTGSEVEMGRPRECL 1011
Db 2792 VSPLSPHRLDTSF-----ASSERLNKAHVSPQ-----KQFIADSTLRLQQT-LPRPMKTL 2840
Qy 1012 GRRSSRIIMETSGMEEBEE-----STAAVPG-----RRGRBD----- 1046
Db 2841 --QKSLDPKPLSPTAESAKERFSLYHQGGGLSQSVSALPNGILVRKVKRTLSPPPPEE 2898
Qy 1047 -----GEVDA-----TASSIPELBRQIE-----KLSKRQ-- 1070
Db 2899 AHLPLAGQVPVQLVAASLLQRLAGPTTVPATKASLLRELDRLRLVEHSTKLKKQAE 2958
Qy 1071 -----LFFRKLLHSSQMLRAVSLGQDRYRRYRWLPPYLAGI-----FVEGTE----- 1113
Db 2959 LDBEKEIDAKLKYLELGIQTKESLAKDRVGRDY--PPLRGLGEHRDYLDSSELNQLR 3015
Qy 1114 --GNLVPVEVIKKTDSLKVAAHSLNLPALFSMKMELAGSNNTTASSPARASRLPKTK-P 1170
Db 3016 LQGTTPAGQVVDVPASAAVPAATPSGPTAFQOPRPPPAATQYTAGSGPTONGFLAQAP 3075
Qy 1171 GFMPORHFKSPVRGQDSQP-----QAQLQPEAQLHVPAQOPQLO-----LQLOSHK 1218
Db 3076 TYTGSTVPAITYPGTSPYAPGPGLPSPQAFPHGTGHYAAPT-PWETTSQAFPPVQADSHA 3134
```

```
Qy 1219 GFLEQEGSPLSLQSGQHDLSQSAFLSMLSTQSHSSLLSSSVLTPDSSP---GKLDPA 1275
Db 3135 AHQKPROTSLA-----DLQ-----KPTNYEVISSPAVTVSTPSETGSGPAVS 3180
Qy 1276 -----QPPEPEPEDEAESPDLOAFWNISQAQPCNAAPTTPPLAVSEDOPTPSPQQLAS 1329
Db 3181 SSYEHGKAPEHPRGDRSS-----VSQSP---APTYP---SDSHVTSLEQNVP 3223
Qy 1330 SKPMNRPSAANPCSPVQFVSTPLAG-----LAPKRAGDPGEMPOSPTGLQP----- 1377
Db 3224 NYVM-----IDDISELTKDSTPTASDSQRPPELPGVSGRPGKDPGEPAVLEGTLPCC 3278
Qy 1378 -----KRRGRPPS--KFFKQMEQRY---LTQLTAQPVVPEM 1408
Db 3279 YGRGEESBEDSYDPRCKSGHRSWESNGRPASTHYYSDDYRHGADKYGPGPMGPK- 3337
Qy 1409 CSGWWITPDPEMLDAMLKALHP-RGIREKALHKLNRHDFLOBSVCLRPSADRIPEPRQL 1467
Db 3338 -----HPSKNLAPAAISSKRSKHK----- 3357
Qy 1468 PAFQEGIMSNSPKE--KTYETDLA 1489
Db 3358 QGMEQKISKFSPIBEAKDVESDLA 3381
```

RESULT 28

```
T47182
hypothetical protein DKFZp434M1616.1 - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 09-Jul-2004
C:Accession: T47182
R:Koehrer, K.; Beyer, A.; Mewes, H.W.; Weil, B.; Wiemann, S.
submitted to the Protein Sequence Database, March 2000
A:Reference number: 224378
A:Accession: T47182
A:Molecule type: mRNA
A:Residues: 1-1299 <AAA>
A:Cross-references: UNIPROT:Q9NSM8; EMBL:AL162004
A:Experimental source: adult testis; clone DKFZp434M1616
C:Genetics:
A:Note: DKFZp434M1616.1
```

```
Query Match 2.48; Score 240; DB 2; Length 1299;
Best Local Similarity 18.64; Pred. No. 0.0033;
Matches 282; Conservative 211; Mismatches 605; Indels 420; Gaps 58;

Qy 38 TNGSPMFPQOGKSLNGDVNVNGLSTVSHTTTSGILNSAPHSSSTSHLHHPVAYDCLWN 97
Db 4 TNGTVNVVAQEPVNTLGDISGNKTPDLISNQNSSDQANEWETASES----- 49
Qy 98 YSQYPSANPGSNLKDPPLLSQ--FSGQYPLNGILGSGRQSPSPSHNTNLRAQSOKFWAN 155
Db 50 -SDPNERRERDEKKNADLNAGTVVKGVE---NVLPPKREIAKRSFSQRPVDRQNRGN 104
Qy 156 GTHSPMGLNFDN--QELYDSFPDQD-----FEEVCSG-----IHPDEAAEKEMTSVVAE 202
Db 105 NGPPKSGRNFSGPRNRRSGPPSKSGKRGPPDDQAGTTGVDLINGSSAHQOE-----GVP 160
Qy 203 NGTGLVCLE-----LLEXOPELKMCGYNGSVPSVSLHQEVSVLVPD--PTVSCLDPP- 254
Db 161 NGTQKNSKDSGTGKKREDPKGPKPEKVDALSQFDLNNYASVVIDDHEVTVIDPQ 220
Qy 255 SHLPD-----QLEDTPILSEDSLEFFN-----SLAPEVSGG 286
Db 221 SNLNDGFTFVWSKKQKRLQDEERRKKEQVIQVWKNKANANEGRSQTSKLPFRFAKKQ 280
Qy 287 LYGIDDTLMEGAEBKPLXOSPVLISALDCSFLNATAFSLIADDSQTSIFASPTSPPV 346
Db 281 ATGIQQAQ---SSASVP-----PLASAPLPPTSSASVASTSAPLATLTPPAS--TSAPV 332
Qy 347 LGES---VLQDNSFDLNGSDAEQEEMETQSSDPPPSLTQAPDQSSSTIQLHPATSPAVS 403
```


706	NLPM-----SCLDSIALGFNFGRHDLSPLASDILNTSGMDEGDDYLPPTTPPAVEKI	759
349	QY-----ESVLQDNSFDLANGSDAEQEEMETQS-SDFP-----PSLTQAPQDSSTLIQLHPA	397
760	PCFPPIESKEBEDKTEQAKVTCGQTTQVETSSSESPFAKEYYKNGTVNAPDLPEMLDLAGT	819
398	TSPAVSPPTSPAVSLVSVPAASPEISBEVCPAASTVVSFAVSVVSPASSAVLPAVSLEV	457
820	RSRLAS-----VSADAEVARARKSVPSAEVVAESSTGLPP-----VADDSQPVKPDQSLE-	868
458	PLTASVTSPKASPYTSPAAAPPTASPANKOVSSPLETTADVEEITGE-GUTASGSGDVMR	516
869	DMGCVFNKTYTLPSP-----VQSENISGSGSFYEGTDDKVR	908
517	RRIATP---EEVRLPLOGHWRREVIKK-----GSHRWQGETWYVPCGRMKQFPPEVIK	568
909	RLDATDLSLIEVKLAAGRVDKETAKEASPPSSADKSGLSREFDQDRANDKLDTV--	966
569	YLSRNLVHSVRREHFSFSPMPVGDFFPEERTDPGLQWVQLSAEBSIPRIOAITGKRGRP	628
967	LEKSEEHVDSKEHAKESEE--VGDKVELFGL--GVTYEQTSAKEL-----ITTKETAP	1015
629	RNTEKAKTKEPKV-----KRGGRPPKVKITELINKTNRPLK-----	667
1016	ERAEKG-LSSVPEVAEVETTKDAQGLDVAAK-----KODQSPLDIKVSDFGQMASGMS	1068
668	-----KLEAQETLNEEDKAK-----IAKSKMKRQKVORCECLTTIQGOARNK	710
1069	VDAGKTYELKEFVQOQLTSLSEAPQETDSFMGISSHVKQAKVSETEV-----K	1118
711	RKQETKSLGHKEAKKXSAEKEBKTK--QEKLK-EKVREK-----KEKYMKKEKE	759
1119	EKVAKPDLVHQEADVKEESYSSGEHSLTWESLKPDEGKKTSPETSLIQDEVALKLSV	1178
760	EVTYAKPACKADKTLATQRRLEERQKQOMLIEHMKKPTEDMCLTDHOPDPFSRVPG-LT	818
1179	EIPCPPVPSEADSS-----IDEAEVQM--EFTQLPKEESTET-----PDIPAIPSDVT	1225
819	LPSGAFSDCLTI VEFLHSFGVLGFDPA-----KDVPSLGVLOEGLCCGDSLGEVQDL	872
1226	QPO-----PNAVSEPAEAVRGEHEETAEAGEYDK-LUFRSDTL-QITDL	1267
873	LVRLKKAALHDPGFPFYCSQSLKILGEKVSEIPLTRDNVSEILRFLMAYGVXPALCDRLR	932
1268	LV-----PG-----SR	1273
933	TOPTQAOPOQKAAVLAPFPHVLANGSTLIINEIDKTLISMSSVRNKNWIVEGRLRUKTV	992
1274	EEFYETCPGEHKGVVSEVVTIEDFITVQTTD-----EGEL-----	1311
993	LAKTRSEVBMGRPEECLGRRSSRIMEETSGMEEBEEESIAAVPGRGRDRGEVDAT	1052
1312	-----GSHSVRFAAPVQ-----PEESRRPYPHDEELEVLMAAQAEPKDGSPAD	1357
1053	ASSIPELERQTEKLSKRQLF--FRKKLLHSSQMLRAVSLGQDRYRRYVWLPVLAGIFVE	1110
1358	AT--PEKEEVPFSEKTYTDYDKDETTIDDSINDADSL-----WV-----DTQ	1399
1111	GTEGNLVPE--EVIKKETDLSUKVAHAASLNPALFSMKMELAGSNWTTASSPARARSPLKT	1168
1400	DDRSIUTEQLETTIPKEERAKEARRPSLE-----KHRRKEKFPKT	1439
1169	KPG-FMOPRH-----FKSPVRQDSEQPOAQI--OPEAQLHVPAQ---QPOLQ	1211
1440	GRGRISTPERREVAKGFSTVSRDEVRRKKAIVYKAEKASEVQAHSPSRKULKPAIK	1499
1212	LQLQSHKGFLEQEGSPLSLQSGOHDLSQASFLMSLQTSQSHSLLSLSSVLTPDSSPGKLD	1271
1500	YTRFTHLSCVCKTKTATIS-GEAQ--APSAF-----KQAKDKYTDGITSPEKRSLSLR	1550
1272	PAPSQPEEPPEDEAGSPDLQAFWNFNSIAQMPCNAAAPTPLPAVSEDOPTSPSQOLASSK	1331
1551	PSSTLPPRRGVSGDREHN-----SESLNSSIARSRTTSPBPIRRAGKSGTSTPTPGS--	1504

RESULT 31

S13507

microtubule-associated protein MAP2 - rat

(for new) and added :bejeds.c

C/Species: Rattus norvegicus (Norway rat)

C;Date: 21-Nov-1993

C;Accession: S13507

R;Marechal, D.; Delapierre, D.; Dresse, A.

Arch. Int. Physiol. Biochim. 96, 231-236, 1988

A;Title: Cloning and partial sequencing of a new rat br

A: Reference number:

A;REFERENCE NUMBER: S13507; MOLD:89334324; PMID:2474284
A:Accession: S13507

A;ACCESSION: SI3507

A;Status: nucleic acid 8

A;Molecule type: mRNA

A;Residues: 1-1825 <MAR>

A;Cross-references: UNIPROT:Q64715; EMBL:X54100; NID:g56624; PIDN:CAA38034.1;

A:Note: the nucleotide sequence was submitted

C:superfamily: microtubule-associated protein MAP2H

Keywords: microbubble binding; tandem repeat

Keywords: Microtubule binding; tandem repeat

Query Match 2 48: Score 238: DB 2: Length 1825:

Query Match: 2.4%; SCORE 238; DB 2
Best Local Similarity: 10.8%; Pred NC: 0.0052

Best Local Similarity 19.8%; Pred: No. 0.0062;

146 RAGSOKFWANG-----THSPMGLNFDSEOELYDSFPDPONFEEVCSGIHP-DEAAEKEMTS 198

QY I48 KASQKFWANG- - - - - IHSFMJLNFDSQEMIDBSIFDQNFEEVCSGJHF - DEARJENMS 139

Q-7

100 IDIA ENCTCI VCCI EI EEYABEI KMCCVNCSEVCEIEST HAEIVSIW VDDPTVSCI DDPSHI.D 258

QY 199 VVAENGTLVCSLELEEXQPELKMCGYNGSVFVESLHQEVSLVLPDFIVSCLDUPSHLP 250

2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525
 526

Qy
259 D-----QLEDTPILSEDSLEPFNSLAPEVPVSGGLYGIDDTELMGAED 300

QY 301 KLPLXDSPVISALDCPSL-----NNATAFSLADDSTST-----IFASPTSPVIG- 348

QV 349 ----ESVLQDNSFDLNNGSDAEQEEMETQS-SDFP-----PSLTQPAPDQSSTIQLHPA 397

.....

398 TSPAVSPTTSPAVSLVSPAASPEISPEVCPAASTVSPA VFSVSPASSAVLPASVLEV 457

[illegible]

Db 865 DMGCVFNKVTPLPSP-----VQDSENLSGSGSYFECTDDKVR 904
Qy 517 RRIATP---BEVRUPLOHGWRREVRIKK-----GSHRWQGETWTYGPCGKMKQFPEVIK 568
Db 905 RDLATDLSLIEVKLAAGRVRKDEFTABKEASPPSADKSGLSRFBFDQDKANDKLDTV-- 962
Qy 569 YLSNLSVHVSRRHFSPRMPVGDFFREERTDPLGLQWVLSABEISRIQAITGKGRP 628
Db 963 -LEKSEBHVOSKEHAKSEB--VGDKVLFGL--GVTYEQTSAKEL-----ITTKETAP 1011
Qy 629 RNTEKATKEVPKV-----KRGGRPPKVKITELLNKTDNRPLK----- 667
Db 1012 ERAEKG-LSSVPEVAEVETTKADQGLDVAK-----KDOQSPLDIKVSDFGQWASGMS 1064
Qy 668 -----KLSAQETLNEDKAK-----IAKSKKQMRQKQVOREGLTTIOGQARNK 710
Db 1065 VDAGKTIELPEVDQQLTLSSEAPQETDSFMGIBSSHVKQGAKVSETEV-----K 1114
Qy 711 RKQETKSLKHKEAKKSKAEKKGTK--QEKLK-EKVKEK-----KEKVMEKE 759
Db 1115 EKVAKPLVQEAVDKESYESSEGESLWESLKPDEGKETSPETSLIQDEVALKLSV 1174
Qy 760 EVTKAKPACKADKTATQRRLEERQKQOMILEEMKKPTEDMCLTDHOPLPDPSRVPG-LT 818
Db 1175 EIPCPVPVSEADSS-----IDKAEVQM-EFQLPKBESTET-----PDIPAIPSDVT 1221
Qy 819 LPSGAFSDCLTIVEFLHSFGKVLGDPDA-----KDVPSGLVQBLEGLCCQSDSGEVDL 872
Db 1222 QPQ-----PEAVVSEPAEVRGESEIEAEGEYDK-LIFRSDTL-QITDL 1263
Qy 873 LVRLKKAALHDPGPSYCQSLKILGKVEIPLTRDNVSEILRCLFMAYGKVPALCDRLR 932
Db 1264 LV-----PG----- 1269
Qy 933 TOPPQAOPPOQAVALAPPVHELNGSTLIINEIDKTLESMSYRKNKWIVEGRRLRLKT 992
Db 1270 BEFVETCPGHEKGVVESVTIEDDFITVQVTTD-----EGEL----- 1307
Qy 993 LAKTGRSEVMGPECLGRRRSRIMEETSGMEEBEEBESIAAVPGRGRRRGEVDAT 1052
Db 1308 -----GSHSVRAFAAPVQ-----PEEERRYPHDELEVLMAEAQAAPKDGSPAP 1353
Qy 1053 ASSIPELERQIEKLSKQELF--FRKLLHSSQMLRAVSLGQDRYRRYVWLPYLAGIFVE 1110
Db 1354 AT--PEKEEVPFYSYKTEYDYDKETDIDSIMDADSL-----WV-----DTQ 1395
Qy 1111 GTEGNLVE--EVIKKETDSLKVAHAASLNPALFMRMELAGSNTTASSPARARSRLKT 1168
Db 1396 DDDRSILTQELETIPKEERAKEARRPSLE-----KHREKEPKT 1435
Qy 1169 KPGFMQPRHFK-----SPVRGDSQPOAQL--QPEAQLHVAQP--OPOLQL 1212
Db 1436 GRGRISTPERKVAKEPSTVSRDVRRAKVAYKKAELAKKSEVQAHSRKLILKPAIKY 1495
Qy 1213 QLQSHKGFLEQSGPLSIGQSHDLQSQAFSLWSLQTSQSHSLSSSVLTPDPSFGKLD 1272
Db 1496 TRPTHLSCVRKKTATS--GESAQ--APSAF-----KQAKDKVTDGITKSEPKSSLRP 1546
Qy 1273 APSQPPEPEDEAESPDQAFWNFISAQMPCNAAPTPPLAVSEDOPTSPQOLAGSKP 1332
Db 1547 SSILPFRGVSGDREEN-----SFLNSGISARRTTRSEPIRRAGKSGTSTPTPGS--- 1599
Qy 1333 MNRPSANPCSPVQFSSTPLAGLAPKRRAGDPGEMPOSPTGLGQPKRRGRPPSKFFQME 1392
Db 1600 ----TAITPGPPSYSS-----RTPGTPG-TPSYRTPPTGTPKFGILVPS----- 1638
Qy 1393 QRYLTQLTQAPVPEMCGWMMI--PDPEMLDMLK-----ALHPRG-----IREKAL 1438
Db 1639 EKKVAIIRTPPKSPATPKQLBLINOPLDLAKNVKSKTGSTDNIKYQPKGGQVIVTKID 1698
Qy 1439 HKHLNKHDRFLQEVCLRPSADPIFEPRQLPAFQEGIMSWSPKETYETDLAVLQWBELE 1498

Db 1699 LSHVTSKCGSLKNIHRPPGGGRV-----KIESVKLDFKEKAQ 1735
Qy 1499 QRVMSDLQIRGWTCPSPDSTRED---LAYCEHLS---DSQEDITWRGPGREGIAPQPKT 1552
Db 1736 AKVGSLD---NAHVPGGNGVKIDSQKLNFRHAKARVDHGABEITITOSPSRSSVASPRRL 1792
Qy 1553 TNPLD-----LAWRLAALAEQNVKRRYLYREPL 1579
Db 1793 SNVSSSGSINLLESPQATLAEDVTAALAKQGL 1825
RESULT 32
A48666
cell proliferation antigen Ki-67, long form - human
C:Species: Homo sapiens (man)
C:Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 09-Jul-2004
R:Schluster, C.; Duchrow, M.; Wohlenberg, C.; Becker, M.H.G.; Key, G.; Flad, H.D.; Gerde
J. Cell Biol. 123, 513-522, 1993
A:Title: The cell proliferation-associated antigen of antibody Ki-67: a very large, ubiq
ins
A:Reference number: A48666; MUID:94043435; PMID:8227122
A:Accession: A48666
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-3256 <SCH>
A:Cross-references: UNIPROT:P46013; EMBL:X65550; NID:g415818; PIDN:CAA46519.1; PID:g4158
C:Keywords: alternative splicing; cell cycle control; nucleus; tandem repeat
F;29-91/Domain: kinase interaction domain homology <KIH>
Query Match 2.4%; Score 237.5; DB 2; Length 3256;
Best Local Similarity 19.2%; Pred. No. 0.014;
Matches 445; Conservative 257; Mismatches 784; Indels 827; Gaps 116;
Qy 1 MEMEANPANDHFNFTGLPPAPAAAGLKPSP---SSGEGL-----YTNQSPMN----- 44
Db 51 IEIHEQEAILH-NFSTNPTQVNGSVIDPEVRLKHGDVITIIDRSFRYENESLQNGRKST 109
Qy 45 -FPQQ-----CKSLNGDVNVNGLSTVSH-TTSGILNSAPH-----SSSTSH 84
Db 110 EFPKRIREQEPARVRSFSSDPDEKAQDSKAYSKITEGKVGNGPNQVHIKNVKEDSTAD 169
Qy 85 LHPSVAYDCLMNVQSPSANPGSNLKP-----PLLSQFSG-GOYPLNGILGG 132
Db 170 DSKDSVAQGTNTVHSSHAGENGNAADPISGDFKEISSVKLVSRYGELKSVPTTQCLDN 229
Qy 133 SRQSPSSHNTNLRAGSQKFWANGTHSPMGLNFDSQELYDSFPDQNFEEVC--SGIHDE 190
Db 230 SKKNESP-----FWKLYESVKKELDVKSQK-----ENVLYQCRKSGLOTDY 270
Qy 191 AAEKEMTSVVAENGTLGVCSLE-----LEEQPELKMCGYNGSVPSVESLHQE 238
Db 271 ATEKESADGL-QGETQLLVSRKSRPKSGGSGHVAEPASPQEQELDQNGKGRDVE----- 324
Qy 239 VSVLVPDPTVSCLDPSHLDPQLEDTP1-LSEDSLEP-----FNSLAPEPV---SG 285
Db 325 -SVUTPSKAVGA-SFPIYEFAMKM-TPVQVSOQQNSPQKHKNKDLTYTGRRESVNLGKSE 381
Qy 286 GLYGIDDT---ELMGAEDKLPL-XDSPVVISALDCPSLNNATAFSLIADDS--QTSTSI 339
Db 382 GFKAGDKTLTPRKLSRTRNTPAKVEDAADSAATKPNLSKTRGSIPTDVEVLPTETET 441
Qy 340 SP-----TSPPVLGESVLQD-----NSFDLNGSDAEQE----- 368
Db 442 EPFLLTLMTQVERKIQKDSLKPEKLTGTAGQCMCSGLPGLSSVDINNPNFGDSINSEGLPL 501
Qy 369 -----EMETQSDPFPPL---TQAPADQSTIQLHPATSPAVPTTSPAVSLVW 414
Db 502 KRRVSPGGHLRPELPDENLPPNTPLKRGAPTREKSLVMH-----TPPVLLKII 551
Qy 415 SPAASP-----EISPEVCPAASTVVSFAVSVSPASSAVLPVAVSLVPLTASVTSP 466
Db 552 KEQPPQSGKQBSGSEIHVEV-KAOSLVISP-----PAPS-----PRKTPVASD 593

[illegible]

QY 1146 ME-----LAGSNTASSPARASRRPLKTPGFQMOPRHFKSPVRGQDSEQ-----1189
 Db 1633 LKTSLGKVGKVEELVAVGKLTQTSGETTHTEPTGDKSMKAFMESPKQLDASAALTG 1692
 QY 1190 POAQLO-PEAQLHVPAPQPOLQ-QLQSH--KGFLQEQSPSLGQSODL-----1237
 Db 1693 SKRQLRTPKGRSEVPDLAGFIELFQTPSHTKSMTNKTKVSYRASQPDLDVDTPTSSK 1752
 QY 1238 -----SOSAFSLWSLQTSQSHSSLLSSSVLTDPDSSPG-----KLDLP 1272
 Db 1753 POPKRSRKAADTEEFLLAFRKQTPS---AGKAMHTPKPAVGEEDINTFLGTPVQKLDQ 1808
 QY 1273 APSOPPE-----EPEDEAESPDLOAFWFNISQAQPCNAAPTPPLVASEDOPTSPPOOLA 1328
 Db 1809 PGNLPGSNRRLOTQRKAQALELTG--FRELFTQCTDNPATADEKTKKILCKSPQSDP 1866
 QY 1329 SSKPMN--RPSAANPCSPVQ-----FSSSTPLAGLAPKRRAGDPGEMPOSPGLGQP-- 1377
 Db 1867 ADPTNTKQRPKRSLLKADVEEBFLAFRKLTTPAGKAMHTPKAAVGEEDINTFVGTPE 1926
 QY 1378 -----KRRGRPPSKFFKMEQ-----RYLTQ-----LTAQPVPPPCWCSGMMW 1414
 Db 1927 KLDLLGNLPGSKRRPQTPKEKAKALEDLAGFKELFTPQGTHTESSMTDDKITEVSKSPQ- 1985
 QY 1415 IPDPEMLDAMLK-----ALHPRGIREKAL-----HKHLNKHRRDFLOE-----VCLRP 1456
 Db 1986 -PDVKTPTSSKQRLKISLGKVGKVEVLVPGKLTQTSCKTQTHRETAGDGKSIAKAKE 2044
 QY 1457 SAUPIPEPRQLPAFOEIGMSW--SPKEKTYETDLAVLQWVEELEQRVMSDLQIRGWTCP 1514
 Db 2045 SAKQMLDPAN--YGTGMEWRPRTPEEA--QSLEDLAGFKELFQ-----2084
 QY 1515 SPDSTREDLAYCBHLSQSDITWRQGREGL----APQRKTNPL---DLAVMRLAALE 1567
 Db 2085 TPDHTBESST-----DBKTKTKIACKSPPPSSMDPTSTRRRPRTPLGKRDI-VEELSALK 2138
 QY 1568 Q-----NVKERYLREPLWPTHEVVLKALLSTPNG-----1597
 Db 2139 QLTQTTHTDKVGDENGINVFRETAKQKLDPAASVTGSKRQRPRTPKGAQPLEDLAGLK 2198
 QY 1598 ----APGTTTETSYETPIRIRIWRQTLQRCS 1626
 Db 2199 ELQTPVCTDKPTTHEKTKI-----ACRS 2223
 RESULT 33
 T13613
 hypothetical protein 8D8.2 - fruit fly (Drosophila melanogaster)
 C:Species: Drosophila melanogaster
 C:Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 09-Jul-2004
 C:Accession: T13613
 R:Papagiannakis, G.; Spanos, L.; Cox, S.; Siden-Kiamos, I.; Louis, C.
 submitted to the EMBL Data Library, April 1999
 A:Description: Sequencing the distal X chromosome of Drosophila melanogaster.
 A:Reference number: Z17695
 A:Accession: T13613
 A:Status: preliminary; translated from GB/EMBL/DDBJ
 A:Molecule type: DNA
 A:Residues: 1-1279 <PAP>
 A:Cross-references: UNIPROT:O46099; EMBL:AL022018; NID:e1273253; PID:e1426350; I
 C:Genetics:
 A:Cross-references: FlyBase:FBgn0024367
 A:Introns: 65/2
 A:Note: EG:8D8.2
 Query Match 2.4%; Score 233.5; DB 2; Length 1279;
 Best Local Similarity 19.3%; Pred. No. 0.0063;
 Matches 294, Conservative 170; Mismatches 521, Indels 537; Gaps 65;
 QY 220 LKMCNGYNGSVPSVE--SLHQEVSVLVPDPFVSCLDPSHLPOOLEPTILSDESLPFFNS 277
 Db 12 LQLGGVTRALPTGGGNAROO-----LDKSFQPPPIEOT-----46

Qy	226	NGSVPSVESLHQ-----EVSVLVDPDPTVS-----CLDDPSHLDPQLEDTP 265
Db	482	VEKLPS--HALHEVVSSANDT7SVSDTKSGLSEDHGVDTNQTIQDCAEELEVVDVN 540
Qy	266	ILSEDSLEPFNSLAPEPVGSLYGIGDTELWGAED--KLPLXO---SPVISALDCPSLNN 320
Db	541	V-----KHAPNEKVGODNSEGNLVNCGDVCLNSAEAEKELPTGDLUSGNASHESAEITLSTNI 596
Qy	321	ATAPSLLDADSO7STSPASPTSPVVLGE--SVLQDNSFDLNNGSDAEQEMETQSSDFF 378
Db	597	DEPLSLL--DTKTAVSDFAE--SSAGVAGEIDAVAME-----SEAAQSIKQCAEAHVA 645
Qy	379	PSLTQAPADQSS7TQLHPATSPAVSP7TSPAVSLVUSPAASPETSPEVCSPAATVSPAV 438
Db	646	PSIIE-----DGEIDRENVCGSEVNVTKTTPAVAREDIPPEKVESEMEESDVKERSINTDEE 702
Qy	439	FSVSPASSAVLPAVSLVPLTASV-----TSPKASPTSPAPAAPPTSPANKOVSSPLET 494
Db	703	VATASVASEIKTCAODLESKV7STD7HTCAKOCVD5QPA-----ENKEGNKLKNEIRL 758
Qy	495	TADVEEITGBGLTA-----SGSGDVMR--RIATPEBVR-LPLOHGWRREVRINKGSHRW 546
Db	759	CTSLVENQKGDVDSIYKLLCGSNVVD7TDKCVASTGEVSVLDASEGLTVAAEIEK-----813
Qy	547	OGETWYGPCGKRMKQPEVIKY-----LSRNLVHSVREHFSPSPMPVGDPEERDTP 602
Db	814	--RPFY-----LPRVPRVDDECLABQLKHABEQ-----VDQK7QNDRALR 852
Qy	603	G-LQWQVLSABEIPSR7QAITGKRGRPRNTEKAKTEKVPKVGGRGPRPPKVKITELINKT 661
Db	853	ADIQIRAIKCDYDISYKAVMAEERSARKAMHSKRQEIALQSMISR---VKSAASVDDI 909
Qy	662	DNRLPKK---LEAQETLNEDKAKIAKSKK-----MRQKVORGEUTTTQOQARNK 710
Db	910	DSRAICKDYDISYKAVMAEERSARKAMHSKRQEIALQSMISRVKSAASVDDIDSRVNM 969
Qy	711	R-----KQETSKSLKH--KEAKKXSAXEKEGKTKQEKLEKVKREKKEKV 753
Db	970	EHTWQHTTSLNBEKGFMREIKQLKUREQISSMGTKDEVK---QALDSKEKTEERLKV 1026
Qy	754	KMKR-----KEEVTKAPACKADKTLATQRRLEE-----RQROQMILEEMKK 795
Db	1027	LRKELDALRNDLSKAEIEITKAAKK-KCDGEWEAQS7KLEQEFRAADA7RQEAFVHLQDLKK 1085
Qy	796	-----PTEDMCLTDHQPLPDFSRVPELTLPSGAFSDCLUTIVE-FLHSF 837
Db	1086	QOREKKNKYFPKYRDNRSAAEMALKDOR-----AALQSLCSDOVENFMNMW 1131
Qy	838	GKVLGF-----DPAKDVP7SLVQLQEGLLCQGD7SLGVEODLLVRL7LKAALHDPGPSPYC 890
Db	1132	NNNDEFKYVYKSN7STR7P7RLGT7-----DGRSLGPDDE-----1166
Qy	891	QSLKILGEKVESEIPL7TRDNVSEILRCFLMAYGVXPALCDRL7RTPFOAQP-----PQOKAA 946
Db	1167	-----PPRITYAPRT-----DKL7RSSDRAEKHEAVPAQKEK 1198
Qy	947	VLAFFVHELNGSTLIIN--EIDK7LESMSYRKKNW7IVE--GR7LRUKTV7LAK7G7SEVE 1003
Db	1199	VWXP-----EGSKVENNGK7E7VAK7PEOK7SOT7TKSKAVK7DQPP7SV7TEL7V--SGKEEIE 1251
Qy	1004	MG-----RPEE7CLGRR7RSR7IMEE7SGMEE7EEB7SIA-AV7PGR7RR7DGEV 1049
Db	1252	KSATP7EE7EP7KL7KEE7EEL7KKEE7EK7KQEA7KMK7EQ7RL7EIA7KAKEM7KK7REE 1311
Qy	1050	DATASSIPELERQIEKLSK7QLPFR7KK7LH7S7QML7RAV7SGQDR 1093
Db	1312	KAKARAVL7KAKAEAEER7EKV7AF7CHSV7OIG--FRSV7SLART7R 1353

RESULT 35

RESOFT 33
T34418

hypothetical protein F12F3.3 - *Caenorhabditis elegans*

C;Species: Caenorhabditis elegans

C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C;Accession: T34418
R;Fulton, B.; Wohldmann, P.
submitted to the EMBL Data Library, July 1998
A;Description: The sequence of *C. elegans* cosmid F12F3.
A;Reference number: Z21521
A;Accession: T34418
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-3488 <FUL>
A;Cross-references: EMBL:U80022; PIDN:AC25985.1; GSPDB:GN00023; CESP:F12F3.3
A;Experimental source: strain Bristol N2; clone F12F3
C;Genetics:
A;Gene: CESP:F12F3.3
A;Map position: 5
A;Introns: 281/3; 332/1; 562/3; 600/3; 1866/3; 1944/3; 3393/1

Query Match	2.3%	Score 230;	DB 2;	Length 3488;
Best Local Similarity	18.5%;	Pred. No. 0.032;		
Matches 326;	Conservative 252;	Mismatches 707;	Indels 480;	Gaps 76;

Qy	166	DSOELYDRSDQNFEEVCSGHP-----DEAAEKMTSVAAE-NGTGLVCSLELEKQP	218
Db	423	DSKSKSDEPEASTEKESTTEKNTDKTSKSAEKKTVKPKKVGTPGLPAKRPVEKKD	482
Qy	219	ELKCMGNGYGVPSVESLHQD---VSVLVPDPTVSCILDPSHLDPQLEDPTILSEDSLEPF	275
Db	483	ASQPSSSKSSPTDGGKKKQIPKALFIPDEISRRFGDPSTMHSETNITITIR-----	535
Qy	276	NSLAPVPSGGLYGIDDTELMGAEADKLPXDSPVISALDCPSLNNATAPSLIADDSQTST	335
Db	536	-----GREG-----SADAKTPLVE-FLSASV-----SMKVFTLVESAKEKAE	571
Qy	336	SIFASPTPPVLGESVLQNSFDLNNGSDAQEE-----METQSSDFPSSLTQAPADQS	389
Db	572	FSFKRSETP-----DDKSRKKEGLPPAKKSEKKDQEVTAEKQSTEALESKKKEVDSE	624
Qy	390	STTQLHPATSPAVSPPTTSPAVSLVSPAAPEISPEV-----CPAASTV--VSPAIVSVV	442
Db	625	KISEQOP-----SDKNKSEVGVPEKAAQFETKQDVSEIEEVPKKTKIKKTEKSDSSI	678
Qy	443	SPASSAVLPVSLVEPLTASVTSKPAKSPVTSAAAAPTASPANK--DVSSFLETTADVEE	500
Db	679	SQKSNVLKPADD--DKSKDDVD-DKSKTTEDQTKVATDSKLEKAADTTTQIETETVDD	736
Qy	501	ITGEGL---TASGSGDVMRRRIATPE--EYRLFLOHGWRREVRIRKGGSHWQGETWYYP	555
Db	737	KSXKKVLKXKTEKSDSFISQKSETPPVPEPTKPAESEAQIAEVNK-----	782
Qy	556	CGRGMKQFPEVIKYVLSRNLVHSVRRHFSPSPMPVGDFFE-----ERDTPEGLQWV	607
Db	783	-AKKQKEVDNLRKREAVAAKTADEKLKTEAENIKKTAEEVAAKQKQKDE-----	834
Qy	608	QLSAE--EIPSRIQAITGKRGRPRNTEKAKTKVEPKVGRGRGRPPKVIITELLNKTDRPL	666
Db	835	QKLETVSVSKSA-----AKLELEKQAQIKAAEADAVKQKELNKNKLEAA	884
Qy	667	KKLEAQETLNEDPAKIAKSKKVRQKVQGECLTTTQOQARNKRQKQETKSLKHKEAKKK	726
Db	885	KKSAADKLLKEESA--AKSKVSESSVKFGE-----EKTKAGEKTVQVSESEPTSKK	935
Qy	727	SXAEEKGKTK--QEKLEKVKKEKKEK-----VKMKKEEVTAKPA--	767
Db	936	TIDTKDVGATEPADETPKKKIIKKKTEKSDSSI--SQKSATDSEKVSQKQODEPT--	993
Qy	768	-----CKADKTLATQRLERQKQOM--ILEEMKPTEDMCLTIDH	805
Db	994	SETQMVTADSKKKQKETDEKLKDAEIAAKTKQEADEKSKLDAQEKIKVSEDDAARKE	1053
Qy	806	QPLPDPSRVPLTLPSPGAFSDCLTIVFLHSFKVLGFDPAKDVPISLGVLOEGLLCQGD	865
Db	1054	KEIENDKLLKLESEIATKASADKLKLE-----QAQAKK	1086


```
Qy 866 LGEVQ-----DLLVRLKAAALHDPGPPSYCQSLKILGEKVSEIPLTRDNVS 911
Db 1087 AAEVEAAKKQEKDEQLKLDTEAASKAAAEKLEKQAQAIKKAAGADAVKKQELDEKN 1146
Qy 912 EILRCFLMAYGVXPALCDRLRTQFPQAPPOQKAAVLAFPPVHELNGSTLIINEIDKTLES 971
Db 1147 KLEANKSAAG-----KLKIEBESAASKQTVEEQ-----KLDAQTKA-----KTAEK 1190
Qy 972 MSSVRK-----NKKVIVGRLLRLKTVLAKTGRSEVEMGRPEBCLGRRRSRMEE 1022
Db 1191 QTKLEKDEKSTKESKETVDKPK--KKVLKKTKESDSSISKSE-----TSKTVE 1242
Qy 1023 TSGMEEBEESIAAVPCRRRRRDEVDATASSIPELERQLEKLSKRQLPFRKKLLHSSQ 1082
Db 1243 SAGSESTQKVA-----ARKQETDEK-----QKLEAEI--TAKKSADKSKLEAESK 1291
Qy 1083 MLRAVSLGQDYYRRYYWVLPYLAGIFVEGTGEGNLVPEEVIKKEKDSLKVAHAHSLNPAIF 1142
Db 1292 LKKAEE-----VEAAKKQEKDQLKLDTEAASKAAAE----- 1325
Qy 1143 SMKVELAGSN--TTASSPARARPLKTKPGFMQPHFKSPVRQDSE----- 1188
Db 1326 --KLELEKQSHIKAAEYADVKKQELBEK-----QRLESEAAATKKADEKLEBEQKKA 1379
Qy 1189 -----QPOALOPEAQLHVPAQ-----POPOLQLQSHKGF 1220
Db 1380 AEIALIETQKEQKLAQEQSLDEAKSAEKQKLESTKSKQTEAPAKESVDKPKKV 1439
Qy 1221 LEQEGSPLSQSHLDLSQSAFLSMLSQTQSHSLSSVL-----TPDSSFG---KLDP 1272
Db 1440 LKKTKESDSSISQKSASAKSTVDAAEATLESDFNLVEKKTQVKQVQSPDESTSATIKRDP 1499
Qy 1273 APSQPPER-----PPDEAESFPDLOAFWNFISAQMPCNAAATPPLAVSEDOPTSPQOLA 1328
Db 1500 A--QKTEISQDDGDEKTTTD-----GKPPKP-----EDSEATPKKRV 1538
Qy 1329 SSKPMNRPSAANPCSPVQFSSTPLAGLAPKRRAGDPGEMPOSPTGLQPKRRGRPPSKFF 1388
Db 1539 KKKTKQSDVASD-----ASLADVSKLDDVE-----EKPKKKVL 1573
Qy 1389 KQEQRYLTQL-----TAQVPPEMCSGWWWTIPDPEMLDAMLKALHPRGIREKALHKH--LN 1443
Db 1574 KKKTEKSDSVISSETSSVDTIKPESVE---IP-TEKAEQMI--LHNRFTSDSAVESBPKN 1626
Qy 1444 XHRPFLQVLCRPSADPIFEPRQIPA-FOEGIMNSPK-----EKYETDLAVLQWVEE 1496
Db 1627 AHKDDTEK-----TTDDMMTRKSSAIFSDDEQSISSKTSSEGRRRRTGFA----- 1674
Qy 1497 LEQRVMSDLQIRGWTCPSPDSTREDLAYCEHLSDSQEDITWRGFGREGLAPQRKTTNPL 1556
Db 1675 --SKFASDTLARG-----DNVIEIELLAEDDTVTWKVNGKDA----- 1711
Qy 1557 DLAVMLAALQONVKRRYLREPLMPTHEVLEKALLSTPENGAPGTTTEISYEITPRIRI 1616
Db 1712 DLN--SRCHEMSHTFTRLIIDEVEPT-----DSGMEITATCGTESHTTILKVEELPVDV 1765
Qy 1617 W---RQTLQRCRAHVCLCLGH--LERS-IAWEK-----SVNKVTCV----- 1654
Db 1766 KYLPKRTSGKEGQEVETISVTLNHPIDISKVYVLKDGKPLEINKOYSIDTVGCVSLTLRR 1825
Qy 1655 CRKGDNDFFLLCGCDRGGCHYCH-RPKMEAVPEGDWFCVCLAAQVGEFTQKPGPK 1713
Db 1826 AKYEDSGKYKVVCDGVDCSTHLSLQGVKLVKNVSETRPVITV-----DKDQFSLLVAYDS 1881
Qy 1714 RGQKRKS----GYSLNFSFEGDGRRR 1734
Db 1882 NPEASFSTVVDGKDLF---DGRSR 1903
```

RESULT 36
A41819
proline-rich peptides 637K precursor, prostatic - rat
C,Species: Rattus norvegicus (Norway rat)

C,Date: 13-May-1992 #sequence revision 13-May-1992 #text_change 09-Jul-2004
C,Accession: A42663; A41819; A31966; B20593; A20593
R,De Clercq, N.; Hemchoote, K.; Devos, A.; Peeters, B.; Heyns, W.; Rombauts, W.
J. Biol. Chem. 267, 9884-9894, 1992
A,Title: The 4.4-kilodalton proline-rich polypeptides of the rat ventral prostate are the 1e exon.
A,Reference number: A42663; MUID:92250652; PMID:1577819
A,Accession: A42663
A,Molecule type: DNA; mRNA
A,Residues: 1-3706; 'I', 3708-4077, 'P', 4079-4155, 'S', 4157-5761 <DE2>
A,Cross-references: UNIPROT:Q63455; GB:M86514
A,Experimental source: ventral prostate
A,Note: sequence inconsistent with the nucleotide translation
A,Note: sequence extracted from NCBI backbone (NCBIN:100347, NCBIp:100348)
R,De Clercq, N.; Hemchoote, K.; Devos, A.; Peeters, B.; Heyns, W.; Rombauts, W.
submitted to the Protein Sequence Database, April 1992
A,Reference number: A41819
A,Accession: A41819
A,Status: not compared with conceptual translation
A,Molecule type: DNA
A,Residues: 1-5762 <DE1>
R,Hemchoote, K.; Peeters, B.; Dirckx, L.; Claessens, F.; De Clercq, N.; Heyns, W.; Wind
J. Biol. Chem. 263, 19159-19165, 1988
A,Title: A single 12.5-kilobase androgen-regulated mRNA encoding multiple proline-rich p
A,Reference number: A31966; MUID:89066721; PMID:3198617
A,Accession: A31966
A,Molecule type: mRNA
A,Residues: 3372-3540 <HEM>
A,Cross-references: GB:M20721; GB:J04188; NID:G206397; PIDN:AAA41950.1; PID:G554494
R,Peeters, B.; Heyns, W.; Bossyns, D.; Rombauts, W.
J. Biol. Chem. 258, 14206-14211, 1983
A,Title: Proline-rich polypeptides bound to rat prostatic binding protein. The primary s
A,Reference number: A94675; MUID:84061859; PMID:6685733
A,Accession: B20593
A,Molecule type: protein
A,Residues: 2020-2057 <PEE>
A,Note: this peptide, designated proline-rich polypeptide V, can be found at several loc
A,Accession: A20593
A,Molecule type: protein
A,Residues: 2822-2859 <PE2>
A,Note: this peptide, designated proline-rich polypeptide IV, can be found at several loc
C,Genetics:
A,Introns: #status absent
A,Note: single copy gene with no introns
C,Superfamily: rat prostatic proline-rich peptides 637K precursor
C,Keywords: prostate; tandem repeat
F,1-26/Domain: signal sequence #status predicted <SIG>
F,27-5761/Product: prostatic proline-rich peptides 637K proprotein #status predicted <MA

Query Match

Best Local Similarity 2.3%; Score 229; DB 2; Length 5762;

Matches 413; Conservative 234; Mismatches 630; Indels 910; Gaps 115;

```
Qy 13 NFTCLPPAPASGLKLP-----SPSSG-----EGLYTN--GSPMFPQQGKSLNG 54
Db 900 SFPVLKPS-AAQALKPRKEKLSPTNNMIPHPSPKPLKNWTHIPAHKWTVPRIQIEDQG 958
Qy 55 DVNVNGLST-----VSHTTSGILNSAPH-----SSSTSHL--HHPS-- 89
Db 959 EYTTISSNGSFQPLDLEVLATSGIIPVKHILPKRTVNPQTYSQVKISHSQHVETQHPNSE 1018
Qy 90 -----VAYDCLWNSQVPSANPGSNLKD-----PP--LLSO---FSGQYPL----- 126
Db 1019 TTTVQPLDLFAINLQPTPKENFAQTQDTTQMIGPPKEVIAQAPHEHGTIPIPIODQ 1078
Qy 127 ---NGILGGRQP-----SSPSHNTNLRAGSKQFWANGTHSPMGLNFDQSLEY 171
Db 1079 AEYTLTVSFQPLDQELTITSEAIRPPHPT---VQQTIIIVHPPEPLVIH--SEQVY 1133
Qy 172 DSFPDQNFEEVCS-----GIHPDEAAEKWTSVVAENGTLGVCSELEEEYQPELKMCGY 225
Db 1134 TQHPNPT-EAIIQPLDLELTITPQTAEGELPQTLQDSTTQII-----EPPTVVV-- 1182
```


QY 226 NGSVP-----SVESLHGVSVLPDPPTVS--CLD-----DPSHLPDQLEDT----- 264
DB 1183 -GPVPIYEVTVQTSQDAEYPPSPTVSFQSLDLELTISPEATRESYHPSLLQOTTIYN 1241
QY 265 -----PILSEDSLEFFN---SLAPEVSGGLYGIDDTLMAEADKLP- 303
DB 1242 PPEHPLVIHSEQVHTQHPNLTEATVQPLDPLTITPQPT-----EGELPQ 1287
QY 304 -LXDSPVISALDCPSLNNATAFSLADDSTSTSFASPTS-----PPVLGESVLDQNSPD 358
DB 1288 TLQDS-----TSQIIIEPDIKVVALPVVQGE-VSQDQA-E 1320
QY 359 LNNGSDAEQEMETQSSDFPSSLTOP-----APQDSSTIOHNPATSPAV----- 402
DB 1321 YTTSSVTSVQPLDQELTITSEAIREFHPTVPOQ--TITVHTKHLVTHSEQTOHPNPT 1378
QY 403 SPTTSP---AVSLWSPAAAPISPEVCPAAASTVSPAFVSVPASSAVLPAVSEVL 459
DB 1379 EVTVQPLDLELMTPOPTAEGELPQTLQDSTTQIIEPPTV-VVGP-----VPI 1425
QY 460 TASVTSPKASPVTSAAAFPTASPANKOVSSFLETTADVEEITGEGLTASGSDVMRRRI 519
DB 1426 YEEVTVQTSQDAEYPPSPTVSFQSLDL-----GLT----- 1457
QY 520 ATPEEVRLPLQHWREVRVTKGSHRWQGETWYGPCGRKMQPFVEVIKLSRLNVHSVR 579
DB 1458 ITPE-----PTEHFITQKTTV-----FPPM--YTDVTLPOQVS 1488
QY 580 REHFSFSPMPVGDFFERD-----TPBG--LOWVQLSAEE----- 613
DB 1489 VOHLK-----PTEGIVQPLDLELTITPQTPPEGELSQTVOESTTQNKEPHKEVAPVPVY 1543
QY 614 -----IPSRIOAITGKRGRPRNTEKAKTKEVPK-----VKRGRGRP 649
DB 1544 QAVTVPTPSQYQAEYQKSLQPLDLELTJTSEPTKEAYHSTISKNSLAINPVVHLOHNP 1603
QY 650 PKVKI-----TELLNKTDNRPLKLE-----AQETLNE----- 677
DB 1604 AEATVQPLDLELTISSSLQPTAEGELLYSMQETVTQISEPPKQVVPVPEYQEVAVPAPV 1663
QY 678 EDKAKIAKSKKMRQKORVGECLTT-IOGARN-KRKOETKSLKHE----- 722
DB 1664 QDAKYPISSIVSLNSLQDELTLSSLELGEALQLTTPDETVMVLPDKRQGIYVDDHDKH 1723
QY 723 -----AKKSAEKEKGTKEKLEKVKREKGVKMKKEEYVTKAP 766
DB 1724 LNLTEVTNPQPHLEHTVQHOPTIEERSQSIQKTTQITEPGCKVVVPLAQSESEVITPWP 1783
QY 767 ACKADKTLATQRRLEERQKQMILEMKKPTE-----DMCLTDHOPLPDFSRVGLT 818
DB 1784 -----ILKETAPPTPHSMALQSLDEKLTIH-----SHSPGWT 1815
QY 819 LPSGAFSDCLTIVEFLHSFGVLGFDPAKVPSLVLQEGLLC-----QG 863
DB 1816 QOHANLKESKG-----HTTGKIL-LDYAEPNMEILKHGLFLKLTTEATTESENTQMT 1869
QY 864 DSLGEVQDILVRLKKAAL-----HDPGPPSYCSQLKILGK-----VSIPLTR 907
DB 1870 KSLKQVTAFTQNKSKMLPALVESQDESPPPNMSLOPLDQELTLSSQPHGHWPHIPNTP 1929
QY 908 DNVSEILRCFLMAYGVXPA-----LCRLRTQPPQ-----AOPQQ 943
DB 1930 EKI-----YLHYAEPTPGFVEPEPPELFFLTKTSRPPVQGTATQMAASPKEWVRAPEN 1982
QY 944 KAAVLAPVHELNGS-----TLIINEIDKTLSSSSYRNKNKI-----VEGR----- 985
DB 1983 KEAVLUSGEGDEQDESPNMSLQSDQEL-TLSS-QPHGWIPHPNTHGKIYLYHYABPP 2040
QY 986 -----LRLKTKVLAK-----RTGRSEVEMGRPEECLGRRRSRKRIMEETSGMBEE 1029
DB 2041 TCPFVEPPDLFLKTKTKSPVWETLRTDKSKEM-----VSQSPKYEAVALPVIGE 2092
QY 1030 EEEESIAAVPGRRRGRDGEVDATASSIPELERQIEKLSKROLFFRKLKLLHSSQMLRAVSL 1089

DB 2093 QOEESRS--PPNMSLQSELEQLTLSSQP-----HG----- 2120
QY 1090 QODRYRRRWV-----LPYL--AGIFVE-----GTEGNLV-- 1117
DB 2121 -----WVHPNTHGKIYLYHYAEPTGTFVEPPDLFLRTTKSKVQGTATRMVKS 2171
QY 1118 PEEVIKTKDSL-----KVAHAASLNALFSMKMELAGSNNTTASSPARASRPJK 1167
DB 2172 PEEVMSLOPENKEAVFPQAQGEKGESPPSNMSLQSLDHELFWMS--SQPHGWIHPHPK 2227
QY 1168 T-----KP--GFMPQ-----RHFKS--PVRGQDSEQQAQLOPEAQLHVPAQOP 1208
DB 2228 TPKKIYLYHYAEPTGTFVEPPDLFLRTTKSKVQGTITTEMAKSPKEMVSO-----TP 2280
QY 1209 QLOLOLQSHKGLEGEPSLSLQSOHDLSSQAFLSWLSQTSQSHSLSSSVLTDPSSSPG 1268
DB 2281 EYKEAVLSPG-EDQDESPSPNTSLKSLDQEVAMS-----SQPH-----SGVHPHPKTPG 2330
QY 1269 K-----LDPAP-----SQP-----PEEPPEDEAESPP 1290
DB 2331 KIYLSHIEPPPGPFVKPDTDLILVKTTKSKPAEWTPRIDKLLKEMVPHSPEVEA--- 2386
QY 1291 DLQATFNISAQMPCNAAPTPPPLAVSEDOPTSPPOOLA-SSKP-----MRPS----- 1337
DB 2387 ---VFPAGHEGQDESGSPNMPL-----QLDQELTLSSQPHGWHVHPHNTPGKIYLYH 2436
QY 1338 -AANPCSPV-----QFSSTPLAGLAKRRAGDPCGM-POSP-----TGLGQ-- 1376
DB 2437 YAPPTGTFVEPPDLFLRTTKSKVQG-SPKELAKSPKEMVSOPEYKEAVLSGPGDQ 2495
QY 1377 -----PKRRGR-----PPSKFFKQMEQORYLT 1397
DB 2496 DESPPPNMSLQSDQEVTMSSQPHSGVPHPKTPGKIYLYHSIEPPPGPFVKPTD-LILV 2554
QY 1398 QLTAPQVPPEMCSGHWIIPDPEDMLDAMLKALHPRGIR-EKALKHKLHNRD----- 1447
DB 2555 KTTTKSKPAE---W--TPRRIDKLLKEMVPHSPYEAEVFPAGHEGQDESGSPPNMP 2606
QY 1448 ---FLOEVCL-----RPSADPIR----- 1463
DB 2607 LQPLDQELTLSSQPHGWHVHPHNTPGKIYLYHYAEPTGTFVEPPDLFLRTPKSPVQGT 2666
QY 1464 PRQLPAFOEGIMSPKPKTY-----ETDLAVLQWVEELQQRVIMSLQIRGWT 1513
DB 2667 PTQMAKSPPEMVSLSPKNKETVFPAGKQDESPSPNLSLQSDQEIETMSS-QPHGW-I 2724
QY 1514 PSPDSTRED--LAYCEHLSDSQEDITWRGREGIAPQKTTNPLDLAVMRLA----- 1564
DB 2725 PHPNTHGKIYLYHAE-----PPTGPFVEPPDLFLKTTKSKPMQ 2765
QY 1565 ---ALEQNVKERYLREP-----LMPPTH 1583
DB 2766 SPRQIDKSPKEMFTQSPYEBSLUPAH 2792

RESULT 37

T42717

DNA-binding protein Rc - mouse

N:Alternate names: Ig kappa chain gene enhancer Recognition component

C:Species: Mus musculus (house mouse)

C>Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 18-Feb-2000

C:Accession: T42717

R:Wu, L.C.; Liu, Y.; Strandmann, J.; Mak, C.H.; Lee, B.; Li, Z.; Yu, C.Y.

Genomics 35, 415-424, 1996

A:Title: The mouse DNA binding protein Rc for the kappa B motif of transcription and for

a family of large transcriptional proteins.

A:Reference number: 222238; MUID:97001141; PMID:8812474

A:Accession: T42717

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-2282 <WUL>

A:Cross-references: EMBL:L46815; NID:g1377885; PID:g1377886; PIDN:AAB40884.1

A; Experimental source: strain BALB/c; clone T1; thymocyte, brain

C; Genetics:

A; Gene: Rc

C; Function:

A; Description: binds V(D)J recombination signal sequence and kappa B motif

C; Superfamily: HIV-EP2 enhancer-binding protein

C; Keywords: DNA recombination; transcription factor

Query Match		2.3%; Score 228; DB 2; Length 2282;
Best Local Similarity		19.5%; Pred. No. 0.023;
Matches		348; Conservative 179; Mismatches 604; Indels 654; Gaps 81;
Qy	17 LPPAPAAAGL-KPSPSSGEGLYTNGSPMN-FPQQKSLNGDVNVNGLSTVSHHTTTSGILN 74	
Db	883 LPPKKRLRLAEMAQSSGESFESVPLSRSPQESSIS-----LS 923	
Qy	75 SAPHSSTSHLHPHSVAVDCLWNTSYQPSANPGSNLKDPLLSQFSGGYPLNGILGSSR 134	
Db	924 GSSRSASFDREDHGKA--EAPGPTSDTRSKTLGSHMLTVP-----SHHPHAREMRRSA 974	
Qy	135 QPSSPS--HNTNLRAGSKFWANGTHSPMGLNFDQSLEYDFFPDQNFEEVCSGIHPDEAA 192	
Db	975 SEQSPNVPHSHMTETRSKFDYGSLSPTGPSL-----AVPAAPPPPPAA 1018	
Qy	193 --EKEMTSVAENGTLVCSLELE----EXOPELKMCGYNGSVPSVSLHGVSVLVDPD 246	
Db	1019 PPERKCFVLVQASLNRRPEALEAVPRENKAVMPA--ASKFSTKSSVQISV-----1071	
Qy	247 TVSCLDLDFSHLPDQLEDPTIILSE-----DSLEPFNSLAPE-PVSGGLYGI-----D 291	
Db	1072 -GTTQGGSPGSKSQMDRPPGLSSPPYTEALQVFLGTQLTPPPASLFLSQLLPQEQQQ 1130	
Qy	292 DTELGAEDKPLXDSPVISALDCPSLNNAFSL-----LAD-DSQ 332	
Db	1131 SSEFFPTQAMAGLSSSPMPPLPPLFPQAPPLPQPTVLHPSQLHQLLPHAADIPFQ 1190	
Qy	333 TSTGIFASPTSP--VLGESVLQDMSFDLNNGSDAEQEMETQSSDRPPSLTOPADQSS 390	
Db	1191 QPPFLNPPCAPSTLSGYFLPQSQFALQIPGEIESHLPPVKYS-LPPLATGPPGSSS 1249	
Qy	391 T-----IQLPATSPAVSTTSPAVSLVVSPPAASPEISPEVCP--AASTVSPSPVSVSP 444	
Db	1250 TEYSSDIQLPVPVQATSP-----APTSAPPLALPACPDAMVSLVVPVRIQTHMP 1299	
Qy	445 ASSAVLPVSLVLEVLPTASVTSPKASPVTSPPAAAPTASPAKDVSSPLETTADVEEITGE 504	
Db	1300 SYGSAMVYTLTSQLVTSQPGSPASTALTK-----YEEPSK---SMTVCEADVYE-----1346	
Qy	505 GLTASGSDVMRRRIATPEEVRLPLQHGMRREVRIKKGSHRWQGETWYVYPCGKRMKQFP 564	
Db	1347 --AEPGPSISKEQ-----NRGYQTP 1365	
Qy	565 EVIKYLSRLNVHVSRRHFSPSPMPVGDFFERTDTPGLOWQLSABEIPSRQAITGK 624	
Db	1366 ---YL-----RVP-----ERKGT-----SLSSGILSLEGCSSTA 1392	
Qy	625 RGRPRNTEKA-----KTKEVPKVKRGRRGPPKVKITTELLNKTNDNRLKLEAQETLINE 677	
Db	1393 SGSKRVLSPAGSLTMTETQOKRVKE-----EASKAD-----EKLELVST---1434	
Qy	678 EDKAKIAKSKKMKROKVORGECLTTTIOGAARKRKQETKS-----LKHKEAKKKXAEK 731	
Db	1435 ---CSVLTSTEDRKTKEPH-----VGGQGRSREAEFLSSLSDDVDSPKLSPLSHSTL 1487	
Qy	732 EKGKTK--OEKLKEVKYKREKKEVKMKKEEYTKAKPACKADKT-LATORRLEERKQKQMI 789	
Db	1488 SHGTAPGSEALKE-----YAPQSSKAHRRGLPPMWSVKKEDPKQETD 1528	
Qy	790 LEEKMKPTEDCLTDHQLPDRFSPVGLTLPFGAFSDCLTIVEFLHSPGKVLGFDPPAKDV 849	
Db	1529 LPPLAPPS-SLPLSDTSP-----KPAK--1549	
Qy	850 PSLGVLOEGLLCOGDSLGEVDQLLVRLLKAAHLDPGPPSYCQSLKILGEVSEIPLTRDN 909	

Db	1550 ----LQEGT----DSKKVLQ-----PPSL-----HTTTN 1570	
Qy	910 VSEILRFLMAYGVKXPALCDRLRTOPFOAQPOQPOQAALVAPPVHELNGSTLIINEIDKTL 969	
Db	1571 VS---WCYLN-----YIKPNHIQCHADRSSYIAG-----1596	
Qy	970 ESMSSYRNKWIIVEGRLLRLKTVLAKRTGRSEVEMRPEECLGRRRRSRIMEETSGBEE 1029	
Db	1597 WCISLYPN-----LPGVSTKAALSLSLK-----QKVSKEYTMTATA 1634	
Qy	1030 EEEESIAAVPGR-RGRRDGEVDATASSIPELER---QIEK-----LSKRQLFF 1073	
Db	1635 PHPEAGRLVPSNKRPMTEVHLPSVSPSQKDPARVEKEKQKAEETPTSKRGSPA 1694	
Qy	1074 RKLLHSSQMLRAVSLQODRYRRRYVLYLAGLFFVEGT-EGNLVPEE--VIKKEITSLK 1130	
Db	1695 RVKIFE-----GGYKSNBEY-----IYVRGRGRGVYCECGIRCKPMSLK 1736	
Qy	1131 --VAHASLNPAL-----FSMKMELAGSNTTASSPARARSPLKTKPGFMQPRHFKSPVR 1183	
Db	1737 KHITHTDVRPYVCKHCHFAFKTK---GNLTKHMKSAHKKCO-ETGVL-----1782	
Qy	1184 GQDSEQQAQIQPEAQHLVPAQPOQLQQLQ-QSHKGFLEQEGSPLSQSQSHDLSQSABL 1243	
Db	1783 ---EELEAEBGTSDLLHQDSEGOEGAEAVEEHQFSDLEDSDSDLDDEDEEEEEE---1835	
Qy	1244 SWLSQTSQSHSLSSSVLTP-----DSSPGKLDPAPOPPPEE-PEPDRAESSPDLOA 1294	
Db	1836 ---EESQDELSGPCSEAAAPCLPPTLOENSPGVEGPOAPDSTDEVPEGSSISEATHLTA 1892	
Qy	1295 FWFNISAQMP-----CNAAPTPLAVSEDOPTPSPOQLASSKP-----1332	
Db	1893 ---SSCSTPSRGTQGLPRGLGLAPLEKDMSSAPSEK-ATSPRRPW-SPSKEAGSRPSVTR 1946	
Qy	1333 ---MNPSPSAANPCSPVQ-----PSTPLAGLAPKR-----1359	
Db	1947 KHSITKNDSSPQQCSAPAREAAQSVTSTPGQMGPRDLGPHLCGSPRPELSRLTPYPIGR 2006	
Qy	1360 -----RAGDPGEMPOSPT---GLQ---PKRRGRP-----PSKFFKQMEQRYL 1396	
Db	2007 EAPAGLERATDTGTTPRYSTRRWSLQGAESPQTVLPFKWALAGPCSPSADKSGLGLQPV 2066	
Qy	1397 TOLTAQVPV-----BEMCSGMMWIPDPEMLDAMLKALHPRGI-----R 1434	
Db	2067 PRALLQVPVLPHTLLSRSPETCTSAWRKTSRSPSAGPGLFPFPFAPHDFHGLHPSRS 2126	
Qy	1435 EKALHKLHKLHRDPL-----QEVCLRPSADPIPEPRQLPAFQEGIMS-----1476	
Db	2127 EENLFSHLPLHSQLSRAPCPLIFIGGIQMVQARPGAQPTVLPGCCAAWVSFGSGGSDL 2186	
Qy	1477 -----WSPKEITYETDLAVLOWVEELEORVIMSDLIQIGWTCPSDPSTREDLAY 1525	
Db	2187 TGAREAQERSQWSPT-----PPDWEFHLRKLQ 2247	
Qy	1526 CEHL---SDSQEDITWPGREGLAGPORKTINPLDLAVMLAALE 1567	
Db	2214 VSKFTLSSELEERTGRGPR-----PPDWEFHLRKLQ 2247	

RESULT 38

149505

adenomatous polyposis coli protein - mouse

N; Alternate names: APC

C; Species: Mus musculus (house mouse)

C; Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 09-Jul-2004

C; Accession: I49505

R; Su, L.

Science 256, 668-670, 1992

A; Title: Multiple intestinal neoplasia caused by a mutation in the murine homolog of the

A; Reference number: I49505; MUID: 92263101; PMID: 1350108

A; Accession: I49505

A; Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: mRNA
A;Residues: 1-2845 <RES>
A;Cross-references: UNIPROT:Q61315; CB:M88127; NID:g19191; PID:g191992
C;Superfamily: adenomatous polyposis coli protein

Query Match 2.3%; Score 227.5; DB 2; Length 2845;
Best Local Similarity 18.9%; Pred. No. 0.032;
Matches 370; Conservative 246; Mismatches 704; Indels 641; Gaps 86;

Qy	29	SPSSGEGLYTNGGPMNPQOGKSLNGDVNVNGLSTVSHTTTSGILNSAPHSSSTSH----	84
Db	1098	SPYRSRG--TSGSETRMGSHAINQNVNQSLOCEDDYEDKPTNYSERYSEREQEHEEE	1155
Qy	85	-----LHHPSVAYDCLWNY-----SOYP-----SANPGSNLKDPLLQSFSG	121
Db	1156	ERPTNYSIKNEKHHVDQIDYSLKATATDISSQKPSFSGKXNSAQSTKPEHLSPSE	1215
Qy	122	GOY--PLNGILGSGRSPSSSHNTNLRAQSKQFWANGTHSPMGLNFDQSBLYSDFPDQNF	179
Db	1216	NTAVPPSNAKRONLRSSAQRNGOTQKGT-----TKVPISNQETIQTYCVEDTPIC	1268
Qy	180	EEVCSGHPDEAAEKEM-----TSVVAENGTGLVCSLEEXQKPELKMCGVNGSVPSVESL	235
Db	1269	FSRCSSLSLSSADDEIGCQDTTQEAEDSANTLQ-TAEVKE-----NDVTRSAEDP	1317
Qy	236	HQEVSVLVPDPTVSCLDPSHLDPQLEDTPILSEDSLE---PNSLAPFVPGGLYGID	291
Db	1318	ATEV-----PAVS--ONARAKPSRLQASGLSSESTRHNKAVFSSGAKSPSKSG----	1364
Qy	292	DTBLMGAEDKLP---LXDSFVSIASLDCPSLNNATAFSLLDADSDTSIFASPTGPPVL	347
Db	1365	-----AQTPSKSPHYHQETPLV-----FSRCTSVSSL--DSFESRSIASSVQSEPCS	1410
Qy	348	GESVLQNSFDLNGSDAQEEMETQSDPPPSLQTQAPQOSSTIQLH-----PATS	399
Db	1411	G-----MVSGIISPDLPSDQGTMPPSRSRKTTPPPQTQAKREVPKSVKVPAAE	1460
Qy	400	PAVSTTSPAVSLVSPASPESPEVCPAASTVVSFAVSVSPAS-SAVLPAVSLVSP	458
Db	1461	KRESGPKQTAVNAQVRV-----QVLDPVDLHLHFATSETPDGFSCSSLALSLEDP	1513
Qy	459	LTA SVTSPKASPVTSAPAAFPFTASPANKOVSSFELETTADVEETIGEGLTAGSGDVMRR	518
Db	1514	FIQKQVELRIMP-----PVQENDNGN-----ETESEQPESSNEN-----QDKE	1551
Qy	519	IATPEEVRLPLQHWREVRV-----KKGSHRWQGETWYVGPCKRMKQFP	564
Db	1552	VEKPDSEKDLDDSDDDDIILEBECIIISAMPTKSSRKAKKLAQTASKLPPVPARKPSQLP	1611
Qy	565	EVTKYL-SRNLVHSVRREHPSFSPRMPVGPFFPEERDTPEGLQWV---QLSAREIISRQ	619
Db	1612	-VYKLLPAQNRLOA--QKHVSFTTGGDDVPVYCVGEGTPINFSTATSLSDUTIIESPNELA	1668
Qy	620	AITGKGRPRNTEKATKEVPKVRGRGRPPVKITELNKTNRPLKLEAQTETNEED	679
Db	1669	TGQGVRAIGSGFEKEDTTPTEGRSTDDAQRKISSIV--TPDLDDNKAEEGDILAECI	1726
Qy	680	KAKIAKSK-----KWRQKVQORGECLTTIQOARN-----KXQBTYSLK-----	719
Db	1727	NSAMPKGSKHKPFVVKIMDPVQOAS--STSGANKNQVDTKEKXKTPSVKPKMPQNTYR	1784
Qy	720	-----HKEAKKSKAEKXGKTQOEL---KEKVK-----	746
Db	1785	TRVRKNTDSKVNTBETPSDNKDSKPSL--QTNAKAFNEKLPNNEDRVRGTFALDSPH	1842
Qy	747	-----REKKEKVKMKKEEVTYKAKPACK-----	769
Db	1843	HYTPIEGTPVCFSRNDSLSLDDFDDVDLSREKAELRKGESKD--SEAKVTCRPEPNS	1901
Qy	770	---ADKLTATORLEERQKQOMILBEMKKTEDMDCLTDHQPLPD-----	810
Db	1902	QQAASKQASIKHPANRAQSKPVLO--KOPTFPQSSKDG---PDRGAATDEKLQNLAIEN	1956

RESULT 39

I38346
elastic titin - human (fragment)

C;Species: Homo sapiens (man)

Qy	811	----	FSRVPGLTIPS	-----	GARSDCLTIVFELHSHFG	-----	KVLGDDP	-----	AKOV	849
			:	:	:	:	:	:	:	
Db	1957	TPVCFSRNSLSLSDIDQENNNNKSEPIKEAEAPANSOQEPKQASGPYAPKSFHVEDT	2016							
Qy	850	P-----	SLGVLEGLLCCGDSILGEVQDILLVLLKAAALHDPGPSPYCQSLKILGEKVSE	902						
			:	:	:	:	:	:	:	
Db	2017	PVCFSRNLSLSL	-----	SIDSEDDLLQCSISSAMPKKRPS	-----					2053
Qy	903	IPUTRONVSEILRCFLMAYGVXPALCDRLTQFPQAPPOOKAAVLAPPVHELNGSTLIH	962							
			:	:	:	:	:	:	:	
Db	2054	-----	RLKSES	-EKQSPKVGILA	-----	EDLTLDL	2079			
Qy	963	NEIDK--	TLESMSYRKY	-----	KWIVEGLRLRLKTVLAKRTGRSEVEMGRPEECGLRRRSS	1017				
Db	2080	KOLQRPDSEHAFSGSEFNFWKAIQEG	-----	ANSTVSLHUAAAAAACLRSQASS	2130					
Qy	1018	----	RIMETS	SG-----	MEEEEBEESIAAVPGRRRRDCEVDATASSIPELERQIEKL	1066				
			:	:	:	:	:	:	:	
Db	2131	DSDSLKLSG	SILSGSPFHLTPDQEEKPFTSNKGPRILKPEKSTLEAKKIESENKGIKG	2190						
Qy	1067	SKQLPFRKKLH-	-----	SSQM-----	LRAVSLGQDRIYRRYVWLPYLAGIFVE	1110				
			:	:	:	:	:	:	:	
Db	2191	GKK----	VYKSLITKIRNSINSEISSOMKQPLPTNMPISIRGRTW	-----	IHIP	2234				
Qy	1111	GTEGNLVP	EEVVIKETDLSLKVAAHAS	-----	LNPALFSMKMELA	1149				
			:	:	:	:	:	:	:	
Db	2235	GLNRSSTSPVSKGPP	LTPASKPSEBQPGATTSPRGTKPAGKSELSP	-ITRQTSQIS	2293					
Qy	1150	GSN-----	TTASSPARARSRPLKTKPGFMQPRHFKSPVR	-----	QGDSEQQAOLQPE	1197				
			:	:	:	:	:	:	:	
Db	2294	GSNKGSSRSGRSDTSPRP	QQPLS	-----	RPQSGPGRNSISPORNGISPPNKLSQL	2345				
Qy	1198	AQLHVA	-----	QPQLOLQLOSHKGFLEQSGPL	-----	SLGSOQ--	1234			
			:	:	:	:	:	:	:	
Db	2346	PRTSSTPASTKSGSGKMSY	TSPGRQLSQONLTQOASLKNASSIPRSESASGKLQMS	2405						
Qy	1235	--	HDLQSAPLWLSOTQSHS	-----	LLSSSVLTPDSSP	--	GKLDP	-----	1272	
			:	:	:	:	:	:	:	
Db	2406	NGNSKNKVELSRN	STKSGSDESSERPALVRQSTFIKEAPSPILRRKLESASFESL	2465						
Qy	1273	APSOPPEEPDPAE	-----	SSPDLQ-----	AFWFNISAQPCNAAPTPIAVSE	1317				
			:	:	:	:	:	:	:	
Db	2466	SPSSRPDS	SPRSQAQTPVLSPSLPDMSLSTHPSVQAGW	----	RKLPPNLSPT	--	TEYND	2519		
Qy	1318	DQPTP-----	SPOQLASSKPMNRPSAANPCSPVQFSSTPLAGLAKPRAGDGPMPQSP	1371						
			:	:	:	:	:	:	:	
Db	2520	GRPTKGRHD	IARSHSESPSRLPINRACGTWKREHSKHSSLP	-RVSTWRRRTGSSSSILSAS	2577					
Qy	1372	TGLGQPKRRGRPS	KPFQMEQRYLTQLTAAQVPPEM	-----	CSGWM	-----	IPDPE	1419		
			:	:	:	:	:	:	:	
Db	2578	S-----	ESSEKAKSEDERHVS	SM--	PAPOMKENOVPTKGTWRKIKESDISPTG	2624				
Qy	1420	MLDAMLKALHPRGIR	EKALHKKHUNKHDFLOEVCILRPSADPIPEPRLPAFQEGIMSWSP	1479						
			:	:	:	:	:	:	:	
Db	2625	MASQSASSGAAS	GAESKPLTIQMAPVSKTDDVWVRIEDCPINNPR	-----	SG	2672				
Qy	1480	KEXTYETDLAVLQW	VELEQRVIMSDLIQIGWTCPSDSTREDLAYCEHLSDSOEDITWR	1539						
			:	:	:	:	:	:	:	
Db	2673	RSPT-----	GNTPPVIDSVSEKGS	-----	SSIKOSKOSKODTH	2704				
Qy	1540	GPREGI	-----	APORKTT-----	NPLDLAVMLAALAEQNVKR	1572				
			:	:	:	:	:	:	:	
Db	2705	GKQSVGSGSPVQ	ITVGLETRLNSFVQVEAFQKGTAKPGOSNPFVISAETAETCIAE	-----	2760					
Qy	1573	RYLREPLWPTHEV	VVLKALLSTNGAPEGTTTISYEITPR	1613						
			:	:	:	:	:	:	:	
Db	2761	--	RTFSSS	-----	SSSKHSSSGTVAARVTFPNTNPSR	2793				

RESULT 39

Db 263 GAVA-----GTRYFCPPKFLGFIAP IHKVIIRIGFPSTSPAKAKTKRMAMGVSA LTHSPS 317
Qy 419 SPEISPEVCPAASTVVS PAVSVVSPASSAVLPVAVLEVLPTASVTSPKASPVTS PAAAF 478
Db 318 SSSISS-----VSSVASVGGPASPASGL-----LTETSRYARKISGTIAL 358
Qy 479 PTASPANKDVSSLETTADVEIITGEGLTASGSDGVMRRRIATPE-----EVRLLPLQHGWR 534
Db 359 QEA-----LKEKQHIQLAE-----RDLEAEVAKATSHICEV-----E 394
Qy 535 REVRIKKGSHRWQGETWYVYCGKMKQFPVVIKYLSNVLVHSVRRHFSPSPMPVGDFF 594
Db 395 KEIALLAKAHE-----QYVAEAEKQLR-----ARLLVENVKKEVDLSNQL----- 437
Qy 595 FEERDTPEGLQWQLSABEISRSQAIT-GKGRPRNTEKAKTKEVPKVKRGRPPKVK 653
Db 438 -EERRKVEDLQF-----RVEE-----ESITKQLETTQTOLEHARIGELEQ----- 476
Qy 654 ITELINKT-D-NRPLKLEAQETLNEEDKAKTAKSKMKRQKVORGE-----CLTTIQGQ 706
Db 477 -SILLEKAAQAE LRELADNRLTTVAEKSRVLQEEEL--SLRGEIEELOHCLLO-SGP 532
Qy 707 ARNRKQBTSLKHKEAKKKSXAEEKGKT-KQBLKEKVKREKKEKVKMEKEEVTAK 765
Db 533 PPADHPAAETLRLRERILLSASKEHQDDSTLLQDKYEHMLKTYQTEVDKLRANEK----- 588
Qy 766 PACKADKTATQRLBERQKQOMILEEMKKTEDMCLTDHQ-PLPDFSRVGLTLPAGAF 824
Db 589 ---YAQEVADILKAKVQOATTENMGLMDNWKSLDLSADHQKSLDLK-----ATLNSQPG 641
Qy 825 SDCLTIVEFLHSFGVLGDFPAKDVPSIGVLQELGCGD-SLGEVDQLLVRLLKALHD 883
Db 642 AQ-----QKEIGELKALVEGIMKEHQLELGNLQ-----AKHD 673
Qy 884 PGFPSYQSLKILGEKVSEIPLTRDNVSEILRCFLMAYGVXPALCDRLRTPFOAQPPQQ 943
Db 674 LETAMHGKEKGLQKQEQVEELAGLQOHWRQLEEQAAASA-----EAQEAQD 723
Qy 944 KAAVLAPFVHLNGSTLIINEIDKTLSEMS---SYRKNWIVEGRRLRLKTVLAKRTGRS 1000
Db 724 QCRDAQLRVQLEGLDVEYRQQAQAEPLKEQISLAEEKMLDYEMLQF-----AEAQSRQ 778
Qy 1001 EVENGREPECLGRR-----SSRIMEETSGMEEBEESIAAVPGRG 1043
Db 779 EAERLEKLLVAENRLOAESLCSAQSHVITESSDLSEETIRM-----KETVEGLQDKLN 833
Qy 1044 RRDGEVDATASSIPELERQIEKLAKRQLFFRKLLHSSQMLRAVSLGQDRYRRRYVWLPY 1103
Db 834 KRDKEVTALTQMDMLRAQVSVLENKCKSGEKKIDSLKERRLEAELEAVSRK----- 887
Qy 1104 LAGFVETGETNLV--PEEVIKKTETDSLKVAHAASLNPALPSMKMELAGSNVTTASSPARA 1161
Db 888 -----THDASQQLVHISQELLRKE-----RSLN-----ELRVLLLEANRHSFGPERD 929
Qy 1162 RSRPLKTPGFWQPHFKSPV-----GODSQPQAQLQPEAQLHVPAPQQLQLOL 1214
Db 930 LSRVHKAEWRIKQKQKLDKIRLEKLTGLDKSEKLSSE-QRRYSLLDPASPPELLKLOH 988
Qy 1215 Q 1215
Db 989 Q 989

RESULT 42
A43427
neurofilament triplet H1 protein - rabbit (fragment)
C/Species: Oryctolagus cuniculus (domestic rabbit)
C/Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 09-Jul-2004
C/Accession: A43427
R/Soppet, D.R.; Beasley, L.L.; Willard, M.B.
J. Biol. Chem. 267, 17354-17361, 1992
A/Title: Evidence for unequal crossing over in the evolution of the neurofilament polype

A/Reference number: A43427; MUID:92381055; PMID:1512270
A/Accession: A43427
A/Molecule type: DNA
A/Residues: 1-606 <SOP>
A/Cross-references: UNIPROT:Q28687; GB:M94315; NID:g164990; PIDN:AAA57152.1; PID:g601930
A/Experimental source: brain
A/Note: sequence extracted from NCBI backbone (NCBIN:112010, NCBI:P:112011)
C/Superfamily: neurofilament triplet H protein
C/Keywords: coiled coil; heterotrimer; intermediate filament; nerve; phosphoprotein
Query Match 2.3%; Score 223; DB 2; Length 606;
Best Local Similarity 23.0%; Pred. No. 0.0073;
Matches 146; Conservative 98; Mismatches 236; Indels 94; Gaps 27;
Qy 189 DEAAEKEMTSSVVAENGTLVCSLEEXQP-----ELKMGYNGSVSVSLEHLQEV 239
Db 31 EGGEEEAASPTBGG---AASPEEEAKSPAEEKSPVKEEAKSPAEEKSPAEEKSPA 87
Qy 240 SVLVDPDPTVSCLDPSHLPDQLEDTPILSEDSLPEFNSLAPPEVSGGLYGIDDTLMGAE 299
Db 88 S---PEKAKSPVKEEAKSPAEEK-SPV-KEEAKSPAEEKSPAEEKS-----PAE 131
Qy 300 DKLP-LKXDPVISALDCPSLNNATAFSLADDSOTSTISFASPTSPVVLGESVLODNGFD 358
Db 132 AKSPEKAKSPVKEEAKSPAEEKSPAEEKSPKAKSP---AEA KSPVKEEAKSP 187
Qy 359 LNNGSDAEQEMETQSSDFPPLTQAPADQDSSITQLHPATSPAVPTTSPAVSLVVS PAA 418
Db 188 EKAKSPVKEEAKSPAEEKSPAEEKSPVKEE-----AKSPEKAKSPAEEKSPA 239
Qy 419 SPEISPEVCPAASTVVS PAVSVVSPASSAVLPVAVLEVLPTASVTSPK--ASPVTSPAA 476
Db 240 SPEKAKS--PEKEEAKSPAEEKSPAEEKSPAEEKSPVKEEAKSPAEEKSPA 297
Qy 477 AFPTA-SPANKDVSSPLETTADVEIITGEGLTASGSDGVMRRRIATPEEVRLLPLQHGWR 535
Db 298 SPEKAKSPVKEEAKSPAEEKSPVKE---EAKSPEKAKSPVKEEAKSPAEEKSPA 354
Qy 536 EVRIKKGSHRWQGETWYVYCGKMKQFPVVIKYLSNVLVHSVRRHFSPSPMPVGDFF 595
Db 355 PEKAK-----SPVKEEAKS-PEKAK--SPEKAKSPVKEE--AKSPEKAKSPVK 397
Qy 596 EERDTPEGLQWQLSABEISRSQAITGKGRPRNTEKAKTKEVP-KVKRGRGPPPKVKI 654
Db 398 EEA KSP-----KAKSPVKEEAKSPEKETPKK-EVKVKEPPKVEETAPPPKV-- 446
Qy 655 TELINKTDNRPLKLEAQETLNEEDKAKI AKSKMKRQKVORGECLTTI QOQARNKRQOE 714
Db 447 -EKDSKKDEAP--KKEAPKPAVEKPKESTAEAKKDEAEDEKKA-APAKMEGKEBAKPK 502
Qy 715 T-----KSLKHKEAKKSAEKECKTKQEKLEKVKREKKEKVKMEKEEVTAKPA 767
Db 503 TEVAKKEPEDAKA KEPSKTEKEPEKPK-KEETPAAPVKEEAKBEARKPEKPKTEAK-A 560
Qy 768 CKADKTLA---TORRLEEROKQOMILEEMKKPTE 798
Db 561 KEDDKALSKPSKPKTEKAEKSSSTQDKDSRPP 594

RESULT 43
G88637
protein F53H1.4 [imported] - Caenorhabditis elegans
C/Species: Caenorhabditis elegans
C/Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 09-Jul-2004
C/Accession: G88637
R/anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A/Title: Genome sequence of the nematode C. elegans: a platform for investigating biolog
A/Reference number: A75000; MUID:99069613; PMID:9851916
A/Note: see websites genome.wustl.edu/gsc/C.elegans/ and www.sanger.ac.uk/Projects/C_ele
A/Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
A/Accession: G88637
A/Status: preliminary

—

—
.
—
.
—
—
—
—

—
.
—
.
—
.
—
—
—
—

Tennant's

www.mindgarden.com

A;Molecule type: DNA
 A;Residues: 1-183,'L',185-969,'N',971-1308,'G',1310-1324,'SS',1326,'HSTLE',1332-1354,'P'
 A;Cross-references: GB:M73548; NID:g190163; PIDN:AA860354.1; PID:g190164
 R;Miki, Y.; Nishishio, I.; Horii, A.; Miyoshi, Y.; Utsunomiya, J.; Kinzler, K.W.; Vogelstein
 Cancer Res. 52, 643-645, 1992
 A;Title: Disruption of the APC gene by a retrotransposon insertion of L1 sequence in a c
 A;Reference number: A44928; MUID:92119623; PMID:1310068
 A;Accession: A44928
 A;Molecule type: DNA
 A;Residues: 1506-1525 <MIK>
 A;Cross-references: GB:S78214; NID:g243541; PIDN:AA821145.1; PID:g243542
 A;Note: sequence extracted from NCBI backbone (NCBIN:78214, NCBI:P:78218)
 R;Spiro, L.; Olschwang, S.; Groden, J.; Robertson, M.; Samowitz, W.; Joslyn, G.; Gelber
 Cell 75, 951-957, 1993
 A;Title: Alleles of the APC gene: an attenuated form of familial polyposis.
 A;Reference number: A49319; MUID:94073973; PMID:8252630
 A;Accession: A49319
 A;Status: preliminary; translated from GB/EMBL/DDBJ
 A;Molecule type: mRNA
 A;Residues: 'G',143-171,'P',173-179 <SPI>
 A;Cross-references: GB:S67787; NID:g461061; PIDN:AAD13997.1; PID:g4261697
 R;Lambertz, S.; Ballhausen, W.G.
 Hum. Genet. 90, 650-652, 1993
 A;Title: Identification of an alternative 5' untranslated region of the adenomatous poly
 A;Reference number: I54271; MUID:93186137; PMID:8383094
 A;Accession: I54271
 A;Status: preliminary; translated from GB/EMBL/DDBJ
 A;Molecule type: mRNA
 A;Residues: 1-4 <LAM>
 A;Cross-references: GB:S56365; NID:g266243; PIDN:AAD14918.1; PID:g4262770
 C;Genetics:
 A;Gene: GDB:APC
 A;Cross-references: GDB:119682; OMIM:175100
 A;Map position: Sq21-Sq22
 A;Note: mutations of this gene can result in familial adenomatous polyposis or sporadic
 C;Superfamily: adenomatous polyposis coli protein
 C;Keywords: cancer; familial adenomatous polyposis; tumor suppressor
 F:1-730/Domain: leucine-rich <NTP>
 F:7-72/Region: coil #status predicted
 F:185-227/Region: coil #status predicted
 F:731-2832/Domain: serine-rich <CTD>
 F:1131-1156/Region: acidic
 F:1558-1577/Region: acidic
 F:1866-1893/Region: highly charged

Query Match 2.3%; Score 223; DB 1; Length 2843;
 Best Local Similarity 18.4%; Pred. No. 0.052;
 Matches 332; Conservative 242; Mismatches 699; Indels 528; Gaps 74;
 QY 75 SAPHSSSTSLHHPSVAVDCLWNYS-QYPSANPGSNLKDPLLQFSGGQYPLNGILGGS 133
 Db 1163 SIKNEEKRHVDQPI-----DYSLKATDIPSSQKQSFSSKSSSQSKTEHMSSSS 1215
 QY 134 RQPSSPS-----HNTNLRAQSQKFWANGTHSPMGLNFDSDLYDSFPDQNFEEVCS 184
 Db 1216 ENTSTPSSNAKQNLHPSSAQRSGQQAATCKVSSINQETQTCYVEDTPICFSCS 1275
 QY 185 GIHPDEAAEKEM-----TSVVAENGTLVCSLELEXPQLKMGYNGSVPSVSHQEVVS 240
 Db 1276 SLSSLSAEDBTGNCQTQAEADSAN-----TLQIAIEIKEI-----GTRSAEDPVSEV- 1323
 QY 241 VLVDPPTVSLDDPSHLPDQLEDTPILSED-----SLRPFNSLAPFVSGGLYG 289
 Db 1334 -----PAVS---QHPTKTSRIQSSLSSESARHKAVFSSGAKGFSKAGQTPKSPPHY 1376
 QY 290 IDDTLMGAEDKPLXDSFVIALDCPSLNNATAPSLIADDSQTSIFASPTSPPVVLGE 349
 Db 1377 VQETPLM-----FSRCTSVSSL--DSFERSIASSVQSEPCSG- 1412
 QY 350 SVLQNSFDLNGSDAEQEEMETQSSDPFSLTOPAPQOSTIQLHPATSPAVSPTTSPA 409
 Db 1413 -----MVSGIISPSDLPSDPCQMTFPPRSKTPPPPPPTAQTAKREVPKNKAPTAEKR 1463

QY 410 VSLVSVSPAASPEIS-PSVCPAASATVSVSPVSVSPAS-SAVLPVAVSLEVP-LTASVTSP 466
 Db 1464 ESGPKQAAVNAQVQVLPDADTLLHFATESTPDGSCSSLSALSILDEPFIOKDVLR 1523
 QY 467 KASPTVSAAAAPPTAS-----PANKDVSSFLETTADVEITGEGLTAGSGGV- 514
 Db 1524 IMPPVQNDNGNTESEQPKESNENQEKAKTIDSEK--DLLDDSDDDDIIELECII 1580
 QY 515 -----MRRRIATPEEV-RLPLQHGWRREVRICKGSHRWQGETWYVGPCCKMKQRPV 566
 Db 1581 SAMPTKSRKAKKPAQTASKLP-----PPVARKPSQLD-V 1614
 QY 567 IKYL-SRNLVHSVRREHFSRPMVQGFPEERTPEGLQWV-----QLSABEIPSRIOAI 621
 Db 1615 YKLLPSQNRLOP--QKHVSFTPGDMPRVICVEGTPINFSTATSLDITBSPNELAAG 1672
 QY 622 TGKGRPRNTEKAKTKEVPKVKRGR-----GRPPKVKITELLNKTNRPLKLAQETLN 676
 Db 1673 EGVRGGAQSGEFKRDITPTGRSTDEAQQGKTSVTIPEL---DDN---KABEGDILA 1725
 QY 677 EEDKAKTAKSK-----KKWRQKVQGECLTTI--QGAENKKEKQETKSL----- 718
 Db 1726 ECINSAMPKGGSHKPFVVKIMDQQAASASSAPNKNQLDGKKKKTPSPVKPIQNTY 1785
 QY 719 -----KHKEAKKKSXAEEKGKTQEKLEKVKREKK-----EKV----- 754
 Db 1786 RTRVRKNADSKNINLAERVSNDKSK-KQNLKNSKDFNDKLPNNEDRVGSGFAFDSPH 1844
 QY 755 -----MKEKEVTKAKPACKADKTLATQRRLEBRQ 784
 Db 1845 HYPIEGTPYCFSRNDSLSLDFDDDDVLSREKAEKAKENKESAKVTSHTLSNQ 1904
 QY 785 ----KQOMILEMKKPTEDMCLTDHQPDPFSRVPGLTLPSGAFSDCLTIVEFLHSGKV 840
 Db 1905 QSANKTOAI---AKQP-----INRGPKPILOKOS--TFPOS----- 1936
 QY 841 LGPDPADKVPVSLGVQLGGLLCQGSGLGEVODLLVRLKAAALHDPFSPYCSQSLKILGEKV 900
 Db 1937 -----SKDIPDRGAATD-----EKLQNFATE-----NTPVCFSHNSLSLSLSDID 1976
 QY 901 SEIPLTRDNVSEILRCLFMAYGVXPALCDRLRTP--FOAQPOOKA---AVLAPPVHEL 955
 Db 1977 QE--NNKNEPIK-----ETEPDQSGEPSPQASGYAPKSFHVEDT 2017
 QY 956 -----NGSTLIINEIDKTLSEMSYRKNKWIPEGRL-----RLKTVLAKRTG 998
 Db 2018 PVSFERNSSLSLSIDSEDDLLQECISSAMPKPKKPKSRLKGNKHSFRNMGILGEDLT 2077
 QY 999 RSEVEMGRPEE-----CLGRRRSS-----RIME 1021
 Db 2078 LDLKDIORPDSEHGLSPDSENFDMKATQEGANSIVSSLHQAAAACLSRQASSDSLS 2137
 QY 1022 ETSG-----MEEEEEESIAAVPGRRRRDGEVDATASSIPELEROIEKLKQRLFF 1073
 Db 2138 LKSGISLSPHLPDQEEKPFTSNKGPRIKPKCEKSTLET-----KKIESEK----- 2186
 QY 1074 RKLLHSSQMLRAVSLGQDRYRRYVWLPYLAGIFVEGTEGVLVPEEVIKETSL---K 1130
 Db 2187 -----GKGG-----KKVYKSLITGK 2202
 QY 1131 VAAHASLNPALFMSKMLAGSNNTTASSPARARSPLTKPGFMQPRHFKSPV--RGQDSE 1188
 Db 2203 VRNSSETSG---QNKQPL-----QANMPSISRGTMTHIFGVRNSSSTSPVSKGPPLK 2254
 QY 1189 QPQAOLQPEAQLHV--PAQPQOLQLOQSHKGFLQEGGSPLSLGQSHDLQSQAFISWL 1246
 Db 2255 TPASKSPSEGQTATTSPRGAKPSVKSEL-----SPVARQTSQIGGSKAPSRSG 2303
 QY 1247 SQTOSHSLSSSVLTTPDSSPGKLDPAQSQ-----PPPEPEDEAESPDLOAFWNISAQ 1302
 Db 2304 SRDSTPSPAQOPLSRIRIQSPGRNSISFGRNGISFPNKLSQLPRTSTASTKSSGSK 2363
 QY 1303 MPCNAAPTPPLAVSEDQTPTS-----PQOLASSKPMNRPSAANPCSPVQFSSTPL 1352

```

Db      2364 M---SYTSPGRQMSQNLTKOTGLSKNASSIPRSESASKGLNMNGAN-----KXVEL 2416
QY      1353 AGLAPKRRAGDPCGEMPSPTGLGQPKRRGRPPSKFFKQMEQRYLTQLTAQPVPPCMCSGW 1412
Db      2417 SRMSSTKSSGSDRSRSPVLVRQ-----STRIKEAPSTLARKLEESASFESLSPS 2468
QY      1413 WNPDPMDLAMLKALHPRGIREKALHKHLNKHDFLOEVCURPSADPIFE-----1463
Db      2469 SRPASPTRSQATPVLSPLSLPDMSLSTHSSVQAGGWRK--LPPNLSPTIEYNDGRPAKR 2525
QY      1464 -----PROLPAFOGINSWPKEKTYETDL-AVLQWVELEQRVMSDLQIRGWT 1512
Db      2526 HDIARSHBESPRKLINRSG--TWKRHSKHSSLPVRVSTW-----RRTGSSSSLSASS 2578
QY      1513 CSPDSTREDLAYCEHLSDSQEDI-----TWRGPGREGIAPQPKTTNPLDLAVMRLA 1564
Db      2579 ESSERAKSEDEKVNISGTSKQSKENQVSAGKTWKIKENBESPNTSISQTVSSGATNGA 2638
QY      1565 -----ALEQNVKRYRLREPLWPHTEVLEKALLSTPNG--APRGTT-----TEISYEITPRI 1614
Db      2639 ESKTLIYQMAPAVSKTEDVW-----VRIEDCPINPRSGRSPGTGNTPPVIDSVSEKANENI 2694
QY      1615 R 1615
Db      2695 K 2695

RESULT 45
B48666
cell proliferation antigen Ki-67, short form - human
C:Species: Homo sapiens (man)
C:Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 15-Mar-2004
C:Accession: B48666
R:Schlueter, C.; Duchrow, M.; Wohlenberg, C.; Becker, M.H.G.; Key, G.; Flad, H.D.; Gerde
J. Cell Biol. 123, 513-522, 1993
A:Title: The cell proliferation-associated antigen of antibody Ki-67: a very large, ubiq
ins.
A:Reference number: A48666; MUID:94043435; PMID:8227122
A:Accession: B48666
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-2897 <SCH>
A:Cross-references: EMBL:X65551
C:Keywords: alternative splicing; cell cycle control; nucleus; tandem repeat
F:2/29-91/Domain: kinase interaction domain homology <KIH>

Query Match      2.2%; Score 220; DB 2; Length 2897;
Best Local Similarity 19.4%; Pred. No. 0.072;
Matches 386; Conservative 223; Mismatches 657; Indels 726; Gaps 101;

QY      204 GTGLVCSLELEEXQPELKMCGYNGSVRSVESLHQEVSVLVPDPTVSCLDPSHLPDQLED 263
Db      30 GRGTECDRIQ-----LPVVSQKHCKIEI-----HSEQEAT-- 59
QY      264 TPILSEDSLEPFNSLAPFVSGGLYGIDDTLMGAEDKLPLXDSFVIALDCPSSLNNA-- 321
Db      60 -----LHNFSSNTPTQVNGSV--IDEPVRLKHGDVITIDRSF--RYENESILONGRK 107
QY      322 -----TAPSLADDSQTSSTIFASPTSPPVLGESVLQDNSFDLNG 362
Db      108 STEFPKRIEQEPARRVRSRFSFSDPEKAQDSKAYSKITEGKVSNGPNELFDENLPPN-- 165
QY      363 SDAEQEEMETQSSDF-----PPSL-----TOPAPDQSSSTIQLHPATSPAVSPITSPAVS 411
Db      166 TPLKRGKAPTARKSLVMHTPPVLKIIKEQOPQSKQESGSEIHVEVK-----AQS 216
QY      412 LVWS-PAASPEITSP-----EVCPAASTVVSFAVSVVSPPASSAVLPAVSLVPLTAS- 462
Db      217 LVISPPAPSPKPTPVASDORRSCKTAPASSKSKQTEVPKGGERVATCLQKRVISERSQ 276
QY      463 -----VTSKPAKSPVTSAPAAFPPTASPAKNDVSSGLETTADVBEI-----TGEGLTASGSG 512

```

```

Db      277 HDILOMICSKRRS-----GASEANLIVA---KSWADVVKLGAKAQOTQTVIKHGQP 323
QY      513 DVMRRRIATBEVRLPLQHGWRREVRICKGSHRWQGETWY-YGPC-----GK----- 558
Db      324 RSMNKRQRRPATPKPV-----GEVHSQFSTGHANSPTCTIIIGKAHTEKVVHP 371
QY      559 -----RMKOPPEVIKYSRLNLVHSVRRE-----HPSFS----- 586
Db      372 ARPYRVLLNFNISQKMDFKEDLSGIAEMFKTPVKEQPOLTSTCHIAISSENLLGKQFQG 431
QY      587 -----PRMPVGD-----FEERD-----TPE----- 602
Db      432 TDSGEELPPTSSEFGNGVFFSAQNAAKQPSDKCSASPLRRQRCIRENGNVAKPRNTYK 491
QY      603 --GLQWQVLSAEIPIRSIQAITGKRGRP--RNTKEK-----AKTKEVPK-----VKRGRG 647
Db      492 MTSLETKTSDETETSPKTVSVNRSGSTERNTOQLPVESKSEBTETIVECILKRGQ- 550
QY      648 RPPKVKITELNKTD-----NRPLKLEAETLINEED-KAKIAKSKKKMRQKVQRGBCL 700
Db      551 -----KATLQORREGEMKEIERPFETVKENIELKENDEKMKAMKRSRTWGQKCAPMSDL 605
QY      701 TTIOG-----QANKRKQETKSLKHKEAKKSYAEKEKG-TK---QEKLEKVKR 747
Db      606 TDLKSLPDTELMKDTARGQNLLOTQD--HAKAPK-----SEKGITKMPCOSLOPEPINT 658
QY      748 EKKEKVKMKE-----KEEV-----TKAKPAKADKTLATQRLREBQK 785
Db      659 PTHTKQQLKASLGKVGKVEALLAVGKTRTSGETTHTHREPAGDG-KSI---RTFKESPK 714
QY      786 QQMILEE-----MKK-----PTEDMLCTDHO----- 806
Db      715 Q--ILDPAARVTGMKKWPRTPKBEAQSLLEDLAGFKELFQTPGPSEE-SMTDEKTTKIACK 771
QY      807 -PLPDFRSVPGLT--LPSGAFSDCLTIVEFL-----HSPQKVL-----GFDPAKDVPS 851
Db      772 SPPSESVDVTSTTQKWPKRSRKADVVEEFALRLKLTSPSAGKAMLTPKPAGD-EKOIKA 830
QY      852 -LGVQLQGLLCCQGSLSGEVODLLVRLKA-ALHD-PGF-----PSYCOSLILGKVKSE 902
Db      831 FMGTVPVKLDLAGTLPKSKQLQTPPKKAKALEDLAGFKELFQTPGHTTEELVAAG-KTTK 889
QY      903 IPL-----TRDNVSEILRCFLMAYGVKPA----- 926
Db      890 IPCDPSQSDPVDPTPTSTKQRPKRSIRKADVEGELLACRLNLMPSAGKAMHTPKPSVGEKD 949
QY      927 -----LCDRLRTQPPQAPPOQQAVALFAPVHLELNGSTLIINEIDKTLSEMS 973
Db      950 IIPFVGTPVQKLDLTENLTGSKRRPQTPKEEAQAL-----EDLTGFKELFQTPGHTTEAVA 1005
QY      974 SYRKNKWIVE-----GRLRLKTLVLAHTGRSEV----- 1002
Db      1006 AGKTTKMPCESSPPESADTPTSTRQPKTPLEKRDVQKLSALKKLTQTSGETTHTDKVP 1065
QY      1003 -----EMGRPEECLGRRRSRRIEETSGMEE----- 1028
Db      1066 GGEDKSINAFREYAKQKLDPAASVTGSKRHPKTKKKAQPLEDLAGWKELFQTPVCTDKPT 1125
QY      1029 -EEEEESIAAVPGRGRDRDGEVDATASSIPELERQIEKLSKRQLFF--RK-----KLLH 1079
Db      1126 THEKTTKIAC-----RSQPD-PVDPTTSSKPSKSLRKVDVEEFFALRKRTPSAGKAMH 1180
QY      1080 SSQMLRAVSLGQDRY-----RRRYWVLPYLAGIFVE--G 1111
Db      1181 TPK--PAVSGEKNIYAFMGTPVQKLDLTENLTGSKRRLQTPKKAQALEDLAG-FKELFQ 1237
QY      1112 TEGNLVPEEVIKKEETDSLKVAHAAS-----LNPAFLFSMKME-----LAGSNWT 1154
Db      1238 TRGH--TEESWTD-KTAKVACKSQPDLDKNPASSKRLKTSLGKVGKVEALLAVGKLT 1294
QY      1155 ASSPARARSRLTKPGFMQFMRHFKSPVRGQDSEQ-----PQAQLQ-PEAQLHVPAQPOPQ 1209
Db      1295 QTSGETTHTEPTGDKSMKAFMESPQKILDSASLUTGSKRQLRTPKGSSEVPEDLAGF 1354

```

QY	1210	LQL-QLQSH--KGFLEQEGSPLSLGOSQHDL-----	1247	-----SQSAFLSWLS
Db	1355	IELFQTPSHTKSMETNEKTTKVSYRASQPDVDTPTSSKPPQKSLRKADTBEFLAPRK	1414	-----
QY	1248	QTSQSHSLSSSVLTDPDSSG-----	1289	-----EPDPDAESS
Db	1415	QTPS-----AGKAMHTPKPAVGEKIDINTFLGTVPVKLDQPGNLPGSNRRRLQTRKEKAQAL	1470	-----
QY	1290	PDLOAFWFNLSAQMPCNAATPPLAVSEDOPTSPQOLASSKPMN---	1346	-----RPSAANPCSPVQ
Db	1471	EELTG--FRELFTQPCDNDPTADEKTTKTLCKSPQSDPADTPTNTKQRPKGLKKADVE	1528	-----
QY	1347	-----FSSTPLAGLAPKRRAGDPGEMPOSPTGLGP-----	1387	-----KRRGRPPSKF
Db	1529	EELFAPRKLTSPAGKAMHTPKAAVGEKIDINTFVGTPEVKLLDGLNLPKSRPQTPKEK	1588	-----
QY	1388	FKQEQ-----RYITQ-----	1430	-----ALHP
Db	1589	AKALEDLAGFKELFQTPGHTESMTDDKITEVSKSPQ--PDPVKPTSSKQRLKISLKG	1646	-----
QY	1431	RGIREKAL-----	1477	-----
Db	1647	VGVKEEVLVPGKLTQTSKGTQTHRETAGDGKSIKAFKESAKQMLDPAN--	1703	-----YGTGKERW
QY	1478	--SPKEKTYETDLAVLQWVEELQORVIMSDLQIRGWTCPSPDSTREDLAYCEHLSQSD	1535	-----
Db	1704	PRTPKEA--QSLDELAGFKELFQ-----	1741	-----DQKTTYK
QY	1536	ITWRGPGREGI-----APQRTKNPL---	1570	-----NV
Db	1742	IACKSPPESSMDTPTSTRRPKTFLGKRDI--VEELGALKQLTQTHTKDVKPGDEDKGINV	1800	-----
QY	1571	KRRYLRPLPWPHEVLEKALLSTPNG-----	1614	-----APEGTTTEISYEITPRI
Db	1801	FRETAQOKLDPAAASVTSKGQRPTRPKGAQPLEDLAGLAKELFQTPVCTDKPTTHEKTTKI	1860	-----
QY	1615	RIWRQTLQRCRS	1626	-----
Db	1861	-----ACRS	1864	-----

RESULT 46

T26517

hypothetical protein Y18D10A.1 - Caenorhabditis elegans

C;Species: Caenorhabditis elegans

C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004

C;Accession: T26517

R;Harris, B.

submitted to the EMBL Data Library, December 1998

A;Reference number: Z20226

A;Accession: T26517

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-1634 <WIL>

A;Cross-references: UNIPROT:Q9XW25; EMBL:AL034393; PIDN:CAA22308.1; CESP:Y18D10A.1

A;Experimental source: clone Y18D10A

C;Genetics:

A;Gene: CESP:Y18D10A.1

A;Introns: 7/3; 44/3; 106/3; 193/1; 500/3; 533/3; 560/3; 590/3; 709/3; 1218/3; 1318/2; 1

Query Match	2.28;	Score	219.5;	DB	2;	Length	1634;
Best Local Similarity	18.98;	Pred. No.	0.037;				
Matches	302;	Conservative	187;	Mismatches	593;	Indels	517;
						Gaps	70;

QY	152	FWANGTHSPMGL-----	NFDSQELYDSPFDONFEVCSCGIHPDEAAEKEMTSVVAENG	204
Db	114	YWRNKYDSVCAITERLEANNIDIKSYE-----	STVKLLETAGKRWTSL--DNQ	160
QY	205	TGLVCSLELEEXOPELKMCCYNGSVPSV-----	ESLHOEVSVL-----	242
Db	161	VTTIKAAKIENSQTKKLDKYLKSVIQVDALSSNTSLSKQSKIKDOLVTVFRDDVRLF	220	

Qy	243	-----VDPPTVSCVL-----DDPSHLDPQL-EDTPIILSDESLPEFNSLAPFEVSG	285
Dd	221	FNSQPWKNSLS*SDPISISATEAEASDPEHLDVDIVTEPAPPIGEQTLS	268
Qy	286	GLYGDIDTELMAEDELPLKXDS-----VISALDCPSLNATAFSLADDQSOTSTIFAS	340
Dd	269	-----DDEE---EEEEEPEADDEVROIENELLOREKVETVRRPDFTAKMIELKA	320
Qy	341	PTSPPVLGVSVL-----QONSFDLNGSDAEOEBEMETOSDPPSPSLTPQAPD	388
Dd	321	KSRAPRLVTETPKVKWPEGHIHKPTEKDSFGLLNSTKIHRNQFFT--SDP-ETTAAQATAER	377
Qy	389	SSTIOLHPATSPAUSPTTSPAVSLVVSPAASEISPEVCPAASTVVSPAVFSVVSPASSA	448
Dd	378	KKALLGAOAGAGASBPGSSSIHGKKGTVR-ELQTEMKGESVKEKAARMRAEASAGRS	436
Qy	449	VLPANSLVPLTASVTSKPASPVTSPAAAAPTASPANKDVSSLETTADVVEITGSLTA	508
Dd	437	QAP-----GPAPAPAASELQDPQDF-----	461
Qy	509	SGSGDVMRRRIATPEBVRPLPLQHGWREVRI-----KKGSHRWQGETWYGPCGKRKM	561
Dd	462	SDPGSDESEEHOKQEKHIPAM-VTRRSARLSALPVTPKKAS-----SSSKMP	508
Qy	562	QPPEVIKYLSRNLVHSVRREHFSPSPMPVGDFEERTDPEGLOWOLSABEIPSRIOAI	621
Dd	509	PPP-----SPSPSTP-----GRRGRPRRTLSTMSMEPAAAAV	540
Qy	622	T-KGRGRPRNTEKAKT-----KEYPK-KVKGRCRP---PKVKITE---LLNKTDNR	664
Dd	541	TPAPRGPRSPSAAKYSENTEPLSEAPSAPVKRGRCRPRSRTSMISTEDSEPSTSSTA	600
Qy	665	PLKLBAEQTLNEEDKAKIAKSKKQROKVQRCELTTIQGOARNRKOBTKSLKHKEAK	724
Dd	601	RSKRASDEE-BEQDLKLTNKPEKPKSKYTE--ETV-GDVLKRLRDT-----AK	649
Qy	725	KKSXAEBKGGTKQEKLKBKRVREKKEKVKMKKEEVTAKPACKADKTLATORRLEERQ	784
Dd	650	TTATVTHTPGPLLTR---KMERMRAPTAVTSSKE-KPKNAGSADSSI---NEEHE	700
Qy	785	KQMILLEMKKPTEDM-----C---LTDHQ-----	806
Dd	701	DETMILEEQ---TLDLPOOTSQOEPRISCGSELLDEQFDASEHSGTVPSAPELTKNPAP	757
Qy	807	EPLPDFSERV---PGLTLPSCARSDCLTIVEFLHSFGKVLGFDAKOVPSLGVLQEGLLCQ	862
Dd	758	PVEASEASAEPKKIDIPQA-TPIIALALALPTVSPT-ALEPPK-----AQEPTAE	808
Qy	863	GDSLGVQDLLVRLKLAALHPGFPSPYCOSLKTLGEKVSEIPLTRDNVSEIL---RCFLM	919
Dd	809	LPTTSISIGRAPQAL-----PTSSQTPPTSG--SAAAPPVDDLSELISGAKTKTKT	856
Qy	920	AYGVXPALCDRLURTOBPQQPOOKA----AVLAFPVHELNGSTLIINEIDKTLESMSY	975
Dd	857	RKAAPPVAVQKSISSTTQQAPPTSVQAPPTSCSAAPPVDDLSELISGAKTKTKTKTMP	916
Qy	976	RKNKWIVEGLRELK---TVLAKTRGREVENGREECJGRRRSSRIMEETSGMBEEEE	1031
Dd	917	PVDOKKISSAEPPIDSAPTSVHQQTPKS-----PQILNISKYGLDISDESEEEEEEE	970
Qy	1032	-----EESIATAVPGRRRRDGEVDATASSIPELERQIEKLSKRQLFFRKXJLHSSOMLR	1085
Dd	971	RGMEIVEEEEEAPPIDSDLQASEPSSTATVKPKVAVVKIPSPE-----	1015
Qy	1086	AVSLGQORRRRRFWJLYLAGIIVFEGTEGNLVPEEVIKKETDLSKVAAHASLNPALFSMK	1145
Dd	1016	-----IDSTSVEAPPEASVPKKA-----PAATKID	1040
Qy	1146	MELAGSNNTTASSPARASRLTKTPGMQPHRFKSPVRGQDSQPQAQLOPEAQLHVPAQ	1205
Dd	1041	NQJLDAQOSEPEPPKARKULFIARK-----IPPKIKISLUPA-	1075
Qy	1206	POPQLOLQSHKGFLEQBSPLSLGCSQHDLSQSFAFLSWLSQTHSSLSLSSVLTPDS	1265

2002 FORD FOCUS

QY 410 VSLVSPASPEISPEVCAAST--VSPAFVSVP-----ASSAVLPAVSEVPLTA 461
Db 1058 M-----PTISTHAPVSPSTPLLTFAANLSPVSVSSSTSLYPDPDPSHK 1112
QY 462 SVT-----SPKASPTSPAAAPTASPANKDVSSFLETTADVEEITGEGLTASG 510
Db 1113 STNWVDVSSQSRHQPK-SPM--PRAQFRSQSP-RHISPTRAITD-----PLAALA 1161
QY 511 SGDVMRRRIATPEEVLPLQHGWRREVRKKGSHRWQGETWYPCGKRMKQFPEVIKYL 570
Db 1162 AAAAARRPIS-----ELPNRIYDSRSKRSRSPRD----- 1193
QY 571 SRNLVHSVRRHFSRPMVGVDFEEDRTPEGLQWVOLSABEIPSRQAITGKGRPRN 630
Db 1194 YRLGTHSLNE-----PVHDYSYDMT-----TPYLGSPRR 1224
QY 631 TEKATKVPKVRGRGPPKVIITELNKTNRPLKLEAQETLNEEDKA-----KIA 684
Db 1225 SRMSTSGERSDDPHHPHSPRGLDFDH-DNQDVROSRATSMLESADVLPQGIKVL 1283
QY 685 KSKKQVRQKVGEGCLTTIQG-ARNKRQETKSLKHKEAKKXSAEKEGKTKOEKLE 743
Db 1284 SPSEHMLPLRGDSTSTESSYGATNVLSATDFIDHSRLERSVTAIEDESSVDSRSKV 1343
QY 744 KVKREKKEKVMK-----EKEVTKAKPACAKADKTLATQRRLEERQKQOMILEMKX 795
Db 1344 KAEELNEDHIPSSTHLVHQNCEPVEVDVPRFCVGTSAEPETALQITQPSHLV----- 1397
QY 796 PTEDCLDTHQPLPFSRVP--GLTLPASGSDCLTIVEFLHSFKVLGFDP-----KDVP 850
Db 1398 PTL-----HPPLASPNELPRSDLSPSTSVVAPLSPLT-----PPATNLKDIP 1440
QY 851 SLGVLEGLLCCQDGLGEVQDLLVRLKAAALHDPGFPSVYCOSLKILGEKVEIPLTRONV 910
Db 1441 PW-----PASPS-----PRNSPSV 1454
QY 911 SE-ILRCFLMAYGVXALCDRLRTOPFOAQPQKAAVLA-----PFVHELN--GSTL 960
Db 1455 SDPVQKSLPPTSLPDAIPSELQPAQVNEARPTLVSADSHAPTDFPMGALSLSVGSSS 1514
QY 961 IINEIDKTLSEMSRKNKIWVGBRLRLKTVLAKTGRSEVEMGRP-EECLGRRRSRI 1019
Db 1515 PIN-ANFTLHAI-----ENK-----ENDCDNECTAQVTRSSPSPMEKILSLPAYAP 1566
QY 1020 MEETSG---MEEEEE-----EBIAAVPGRRRRDEVDATASSTPELER 1061
Db 1567 IEQSSSGPVHRETEVNCPPNTTHIIPSELPAASVGSFQQDPPLHASIHDVPALE- 1625
QY 1062 QIEKLSKRQLFPRKKLHSSQMLRAVSLGQDRYRRRYWLPYLAGIFVEGTEGNLPVEV 1121
Db 1626 -----HSSM-----SPLHSPGNDSPPTPKEV 1647
QY 1122 IKKETDSLKV-----AAHASINPALFSKMWELAGSNTTAS-- 1156
Db 1648 DSEEVETMDVDVWNAHDTNDQERKALQTLIPSTINTNLSPLVSVRAEINSEHIDASEL 1707
QY 1157 ---SPARAS-----RPLKTKPGFWQPRHFKSPVRGQDSEOP- 1190
Db 1708 APQTSPALATSPPTTIFAASKVKMSLKDFAIRKRQKEEBEREKKEQDKGNDVQKP 1767
QY 1191 -----QAQLOPEAQLHVP-----AQOQFO-----LQLOQSHK 1218
Db 1768 VDWEADAIIVEDRQKVQESHVEAPVDSNGEDDAEKPETDNLSPAPPKYVVPVLESN- 1826
QY 1219 GFLEQEGSP-----LSLQSQSHDLSQSAFLSWLSQTSQSHS----- 1253
Db 1827 --IEERGHVAKSDVDVIMKASETLNKLINSSMSMKMTVAEYASGPPHFDVSTAKL 1884
QY 1254 -----SLSSSVLTPDSSPGKL-----DPASQPPEEPEDAEAS-- 1288
Db 1885 PPMVSLSASPNIASRPRGLITDPIIADSSLSLLREKTMAPHSPVEDGEDIGDGR 1944
QY 1289 -----SPDLQAFWFNISAQMPCNAAPTPLA-----VSEDOPTSPQOLASS 1330

Db 1945 MFRFSPKPHSPSNIKSYL--LSTSQPKSYSP-PPRAPKIHENITNEQRKTPPTQPRS 2002
QY 1331 KPMNRPSAANPCSPVQFSSTPLA-----GLAPKERAG-----DPGEMQSP 1371
Db 2003 FTASPSV---SSVPHGSPPIRLALAPVANGLSPKVAANVTSTSSAPFIPSPASAN 2058
QY 1372 TGLQPKRRGRHPPS--KPFKQMEQ 1393
Db 2059 PALGR-----TPPSGRALRLOQ 2077
RESULT 50
150591
class II INCENP protein - chicken
C:Species: Gallus gallus (chicken)
C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 09-Jul-2004
C:Accession: 150591
R:MacKay, A.M.; Eckley, D.M.; Chue, C.; Earnshaw, W.C.
J. Cell Biol. 123, 373-385, 1993
A:Title: Molecular analysis of the INCENPs (inner centromere proteins): separate domains
A:Reference number: A48482; MUID:94012983; PMID:8408220
A:Accession: 150591
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-877 <MAC>
A:Cross-references: UNIPROT:P53352; EMBL:Z25420; NID:G414110; PIDN:CAA80907.1; PID:G41411
Query Match 2.2%; Score 217; DB 2; Length 877;
Best Local Similarity 21.3%; Pred. No. 0.022;
Matches 179; Conservative 112; Mismatches 313; Indels 238; Gaps 39;
QY 93 DCLW-----NYSQPSANPGS-----NLKDPPLLSQFSGQYPLNGILGSR 134
Db 29 DLLWLEIEBAARMFGSHYDQELMPKTSQKRRKRKPSALRGESLEL-----GR 82
QY 135 QPSSPSHNTNLRAGSKFWANGTHSPMGLNFDISOELYDFFDQNFEEYCSGHPDEAAEK 194
Db 83 RRLSRRRTNNLKAVSSK-----RDSQRL-----QNKEDT-EGLGTE-AQE 120
QY 195 EMTSVVAENG-GLVCSLEEXQPELMCMGNGSVSPVESLHQEVSLVPDPTVSCLD 252
Db 121 LSSQTVSRRLTRSOVAAPADRSEVLP-----EHLRERV---VPVVEISVCD 163
QY 253 DPSHLPOLETPILSEDSLEFPNSLAPPEVSGGLYGIDDTLMGAEDKLPLKDS----- 307
Db 164 ---RISAEFQKCASERAENHSASLPP-----SSDDKSPKESAAESQ 204
QY 308 --PVISALDCPSL-----NNATAFLLADDSQTSIFASPTSPVVLGESVLDQNSFD 358
Db 205 PLPAASELIVPHTPEAKGAGKNSAFKKTANVAUTV-----VLSEKEIG 249
QY 359 LNNSGDAQEEMETQSSDFPS-ITQAPADQSSITQLHPAT-----SPAVSPTTSPAV 410
Db 250 LEEVDSDAQVKHNERDDKEPSQRTDTPETPTGSLRRSRVRRSLMGKPSITIRTSLAE 309
QY 411 SLVSVPAASPISEVCPAASVTVSPVSVVSPASSAVLPAVSLVP---LTASVTSPK 467
Db 310 KYSLARKRESTIRKSI---ARTVTKRKAQKLSVSSSVNGSGSEVEPDEBETVTVNAGPP 366
QY 468 ASPTVSPAAAPTASPANKDVSSFLETTADVEEITGEGLTASGSDVMRRR----- 518
Db 367 PVQTPPKLDF-----QGLRMSLSRQTVNRNEQQQETSNN 401
QY 519 ---IATBEVRLPLQHGWRREVRKKG-SHRW---QGETWYPCGKRMKQPP-----EVI 567
Db 402 ECDLSKSEKTOEPQSA--RRKTSYKRAVDQYDQAEQDGGLSPLRKKTPTSPPCPASKV 460
QY 568 KYLSRLNVHSVRRHFSRPMVGVDFEEDRTPEGLQWVOLSABEIPSRQAITG 623
Db 461 RPF-KTFLHTVEKNQLLMTSPSSVGRNGVIKFKYKNTP--LQHDPKKER--QKLAARK 515

QY

624

KGRPRNTEKATKEVPKVKRGRGRPPKVKITELLNKTDNRPLKKLEAQETLNBEEDKAKI

683

Db

516

K-----EEAEQLRKQKVEBEKKRQBEAKLRREERLRKV-----LQARERAEQLEEEERKRI

567

QY

684

-----AKSKKMR---OKVORGECLTTIQOARNKR--KQETKSLK

719

Db

568

BQKLALFDEKTEKAREERLAEKIKGAAAKMEEAEARRRQDBEARKQKALQOQEEERR

627

QY

720

HKEAKKKSXAEK-----EKGKTQBEKLK-----EKVKREKKEKVKMKEKEEVTAKPA

767

Db

628

HKELMQKKKEEEOEFARKIADQRAEQEREKQLAAERQERKKQBERKKEEERIQA EKQR

687

QY

768

CKADK-----TLAT-----QRRLEERQKQOMILEEMKK-PTEDWCLTDHQPLPDFSRVP

815

Db

688

EQQEKAAARLQKEVLAAKEQLQKEMEKEKEKEQLLAEMKRRQEQEKCLPEEQKAKDVACTQ

747

QY

816

GL 817

Db

748

HL 749

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 9, 2005, 14:24:09 ; Search time 103.292 Seconds
(without alignments)
4877.138 Million cell updates/sec

Title: US-10-702-148-29
Perfect score: 7985
Sequence: 1 NAPLGRKPFPLVNPLOEE.....NCFMMLVNTQFCMALTDITVT 1531

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1391452 seqs, 329044822 residues

Total number of hits satisfying chosen parameters: 1391452

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 65 summaries

Database : Published Applications AA:*

1:	/cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2:	/cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
3:	/cgn2_6/ptodata/1/pubpaa/US05_PUBCOMB.pep.*
4:	/cgn2_6/ptodata/1/pubpaa/US04_PUBCOMB.pep.*
5:	/cgn2_6/ptodata/1/pubpaa/US03_PUBCOMB.pep.*
6:	/cgn2_6/ptodata/1/pubpaa/US02_PUBCOMB.pep.*
7:	/cgn2_6/ptodata/1/pubpaa/US01_PUBCOMB.pep.*
8:	/cgn2_6/ptodata/1/pubpaa/US00_PUBCOMB.pep.*
9:	/cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
10:	/cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
11:	/cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
12:	/cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
13:	/cgn2_6/ptodata/1/pubpaa/US05_PUBCOMB.pep.*
14:	/cgn2_6/ptodata/1/pubpaa/US04_PUBCOMB.pep.*
15:	/cgn2_6/ptodata/1/pubpaa/US03_PUBCOMB.pep.*
16:	/cgn2_6/ptodata/1/pubpaa/US02_PUBCOMB.pep.*
17:	/cgn2_6/ptodata/1/pubpaa/US01_PUBCOMB.pep.*
18:	/cgn2_6/ptodata/1/pubpaa/US00_PUBCOMB.pep.*
19:	/cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
20:	/cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	7985	100.0	1531	9	US-09-839-479-29
2	7985	100.0	1531	15	US-10-376-537-29
3	7985	100.0	1531	15	US-10-702-148-29
4	7955	99.6	1527	9	US-09-839-479-27
5	7955	99.6	1527	15	US-10-376-537-27
6	7955	99.6	1527	15	US-10-702-148-27
7	7876	98.6	1525	9	US-09-839-479-68
8	7876	98.6	1525	15	US-10-376-537-69
9	7876	98.6	1525	15	US-10-702-148-68
10	2877	36.0	572	9	US-09-864-761-34546
11	2835	35.5	560	9	US-09-864-761-37938
12	930.5	11.7	1674	9	US-09-839-479-1
13	930.5	11.7	1674	15	US-10-376-537-1

ALIGNMENTS

RESULT 1

US-09-839-479-29
; Sequence 29, Application US/09839479
; Publication No. US20020039779A1
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042002
; CURRENT APPLICATION NUMBER: US/09/839,479
; PRIOR FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24

14	930.5	11.7	1674	15	US-10-702-148-1	Sequence 1, Appli
15	919	11.5	1673	9	US-09-839-479-69	Sequence 69, Appl
16	919	11.5	1673	15	US-10-376-537-70	Sequence 70, Appl
17	919	11.5	1673	15	US-10-702-148-69	Sequence 69, Appl
18	712	8.9	141	9	US-09-764-864-1134	Sequence 1134, Ap
19	599	7.5	1586	13	US-10-087-192-663	Sequence 663, App
20	533	6.7	1972	9	US-09-839-479-21	Sequence 21, Appl
21	533	6.7	1972	15	US-10-376-537-21	Sequence 21, Appl
22	533	6.7	1972	15	US-10-702-148-21	Sequence 21, Appl
23	531.5	6.7	1873	13	US-10-087-192-666	Sequence 666, App
24	518.5	6.5	1876	9	US-09-839-479-70	Sequence 70, Appl
25	518.5	6.5	1876	15	US-10-376-537-71	Sequence 71, Appl
26	518.5	6.5	1876	15	US-10-702-148-71	Sequence 71, Appl
27	512.5	6.4	1878	9	US-09-839-479-13	Sequence 13, Appl
28	512.5	6.4	1878	15	US-10-376-537-13	Sequence 13, Appl
29	512.5	6.4	1878	15	US-10-702-148-13	Sequence 13, Appl
30	509.5	6.4	1969	9	US-09-839-479-71	Sequence 71, Appl
31	509.5	6.4	1969	15	US-10-376-537-72	Sequence 72, Appl
32	509.5	6.4	1969	15	US-10-702-148-71	Sequence 71, Appl
33	341	4.3	65	9	US-09-864-761-34543	Sequence 34543, A
34	338	4.2	69	9	US-09-864-761-34544	Sequence 34544, A
35	322	4.0	1790	15	US-10-369-493-1586	Sequence 1586, Ap
36	311	3.9	65	9	US-09-839-479-67	Sequence 67, Appl
37	311	3.9	65	15	US-10-376-537-68	Sequence 68, Appl
38	311	3.9	65	15	US-10-702-148-67	Sequence 67, Appl
39	305.5	3.8	1327	16	US-10-408-765A-1215	Sequence 1215, Ap
40	302.5	3.8	3225	16	US-10-408-765A-254	Sequence 254, App
41	299	3.7	2375	16	US-10-408-765A-277	Sequence 277, App
42	297	3.7	1400	9	US-09-764-176-7	Sequence 7, Appli
43	296	3.7	2492	16	US-10-697-526-2	Sequence 2, Appli
44	294	3.7	1398	15	US-10-094-466-30	Sequence 30, Appl
45	293	3.7	1881	14	US-10-032-585-7646	Sequence 7646, Ap
46	290.5	3.6	1972	14	US-10-171-311-162	Sequence 162, App
47	290.5	3.6	1972	15	US-10-341-434-103	Sequence 103, App
48	290.5	3.6	1979	10	US-09-927-537-4	Sequence 4, Appli
49	290	3.6	1827	15	US-10-369-493-5368	Sequence 5368, Ap
50	284	3.6	58	14	US-10-029-386-30379	Sequence 30379, A
51	279	3.5	1819	15	US-10-335-977-7981	Sequence 7981, Ap
52	276.5	3.5	2543	16	US-10-828-985A-9	Sequence 9, Appli
53	276	3.5	1938	14	US-10-171-311-164	Sequence 164, App
54	276	3.5	1945	10	US-09-927-537-2	Sequence 2, Appli
55	275.5	3.5	2020	15	US-10-369-493-5128	Sequence 5128, Ap
56	275.5	3.5	2020	15	US-10-369-493-5129	Sequence 5129, Ap
57	275	3.4	1820	15	US-10-335-977-7982	Sequence 7982, Ap
58	275	3.4	1927	15	US-10-282-122A-58729	Sequence 58729, A
59	274.5	3.4	2503	16	US-10-828-985A-11	Sequence 11, Appl
60	274.5	3.4	2568	16	US-10-828-985A-7	Sequence 7, Appli
61	272	3.4	149	14	US-10-195-730-317	Sequence 317, App
62	272	3.4	149	16	US-10-799-747-317	Sequence 317, App
63	270	3.4	50	9	US-09-864-761-40113	Sequence 40113, A
64	266.5	3.3	1879	15	US-10-296-115-1265	Sequence 1265, Ap
65	265	3.3	59	9	US-09-839-479-66	Sequence 66, Appl

		Best Local Similarity 100.0%; Pred. No. 0;		Matches 1531; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
Qy	1	MAPLGRKPPPLVNPPLGPEPPFTI	PHTQEA	PTREEYE	EARLERYSERIWTCKSGSSQL 60
Db	1	MAPLGRKPPPLVNPPLGPEPPFTI	PHTQEA	FTREEYE	EARLERYSERIWTCKSGSSQL 60
Qy	61	THKEAWEEQVVAELLKBEFFPAWYK	KL	VLEMHVHNTAS	LEKLVDTAWLEIMTKYAVGEEC 120
Db	61	THKEAWEEQVVAELLKBEFFPAWYK	KL	VLEMHVHNTAS	LEKLVDTAWLEIMTKYAVGEEC 120
Qy	121	DFEVGKMKLVKIKVHLEKVDDEATE	KKSDGACD	SPSSDKENSQ	IAQDHQKETVY 180
Db	121	DFEVGKMKLVKIKVHLEKVDDEATE	KKSDGACD	SPSSDKENSQ	IAQDHQKETVY 180
Qy	181	KEDEGRRESINDRARRSPKLP	TS	LKGERKWA	PKFPHKYDVKLQNEDKIISNVPADS 240
Db	181	KEDEGRRESINDRARRSPKLP	TS	LKGERKWA	PKFPHKYDVKLQNEDKIISNVPADS 240
Qy	241	LIRTERPPNKEIVRYFIRHNAL	RAGTG	GENAPW	VEDELVKYSLPSKPSDFLLDPYKYMT 300
Db	241	LIRTERPPNKEIVRYFIRHNAL	RAGTG	GENAPW	VEDELVKYSLPSKPSDFLLDPYKYMT 300
Qy	301	LNPSYKRYKOTGSPDRKPSK	KTDSN	SSLSPLNPKL	WCHVHLKSLSGSPLYKQNSKNSK 360
Db	301	LNPSYKRYKOTGSPDRKPSK	KTDSN	SSLSPLNPKL	WCHVHLKSLSGSPLYKQNSKNSK 360
Qy	361	SPEEHLEEMKQMSPNKLHNT	FHPIPKG	PPAKPGKHS	DKPLKAKGRSGIILNGOKSTGN 420
Db	361	SPEEHLEEMKQMSPNKLHNT	FHPIPKG	PPAKPGKHS	DKPLKAKGRSGIILNGOKSTGN 420
Qy	421	SKSPKKGLTKPTKMKQMTLL	DMAKGTQ	QWTRAP	NSGGTPTSSKPKHKLPPAALHLIA 480
Db	421	SKSPKKGLTKPTKMKQMTLL	DMAKGTQ	QWTRAP	NSGGTPTSSKPKHKLPPAALHLIA 480
Qy	481	YKKNKDRDKSALS	CVISKTAR	LSSDRARL	PEELSLVQKRYELLEHKKRWASME 540
Db	481	YKKNKDRDKSALS	CVISKTAR	LSSDRARL	PEELSLVQKRYELLEHKKRWASME 540
Qy	541	EORKEYLKKREBELKKLKE	KAKERRE	KEMLEKQRYE	QDELTKGNLPAFRLVDTPTE 600
Db	541	EORKEYLKKREBELKKLKE	KAKERRE	KEMLEKQRYE	QDELTKGNLPAFRLVDTPTE 600
Qy	601	GLPNTLFGDVAMVVEFL	SCYSGLL	PDQAPITAV	SLMEALSADKGGFLYLNRLVILIQ 660
Db	601	GLPNTLFGDVAMVVEFL	SCYSGLL	PDQAPITAV	SLMEALSADKGGFLYLNRLVILIQ 660
Qy	661	TLLQTLLODEIAEDYGEL	GMKLS	SEIPLTH	SVSELVRLCLRRSDVQEESEGSDTDDNKDS 720
Db	661	TLLQTLLODEIAEDYGEL	GMKLS	SEIPLTH	SVSELVRLCLRRSDVQEESEGSDTDDNKDS 720
Qy	721	AAPEONEVODEFLEK	LETSEFF	ELTSEELQ	ILTALCHRLIMTYSVQDHMETRQOQSABL 780
Db	721	AAPEONEVODEFLEK	LETSEFF	ELTSEELQ	ILTALCHRLIMTYSVQDHMETRQOQSABL 780
Qy	781	WKERLAVLKEENDKKRAE	QKQKEMAK	KNKENG	VENGLGKTDKRIIVKFEPOVDTEAE 840
Db	781	WKERLAVLKEENDKKRAE	QKQKEMAK	KNKENG	VENGLGKTDKRIIVKFEPOVDTEAE 840
Qy	841	DMISAVKSRLLAIQA	KKREIIOER	MKVLERQAE	EERIRKHAAAEKAFQEGIAKAKL 900
Db	841	DMISAVKSRLLAIQA	KKREIIOER	MKVLERQAE	EERIRKHAAAEKAFQEGIAKAKL 900
Qy	901	VNRRPTIGTDRNHRN	YWLPS	DEVPLF	IEKGWVHDSIDYRFNHCKDHTVSGDEYCPRS 960
Db	901	VNRRPTIGTDRNHRN	YWLPS	DEVPLF	IEKGWVHDSIDYRFNHCKDHTVSGDEYCPRS 960
Qy	961	KCANLGKNA	SMNTQHG	TATEVAVET	TPKQGNLWFLCDSQKXELDELLNCLHPQGIRESQ 1020
Db	961	KCANLGKNA	SMNTQHG	TATEVAVET	TPKQGNLWFLCDSQKXELDELLNCLHPQGIRESQ 1020
Qy	1021	LKERLEKRYQDIIH	SIHLARK	PNLGLK	SCDGNQELLNFLRSLIEVATRLQKGLGYVEE 1080

Db	1021	LKERLEKRYQDIIH	SIHLARK	PNLGLK	SCDGNQELLNFLRSLIEVATRLQKGLGYVEE 1080
Qy	1081	TSEFEARVISLEK	LQDFGEC	VIALQAS	SVIKKFLQGFMAPQKRRKLQSDSDSAKTEEVDEE 1140
Db	1081	TSEFEARVISLEK	LQDFGEC	VIALQAS	SVIKKFLQGFMAPQKRRKLQSDSDSAKTEEVDEE 1140
Qy	1141	KQWBEAKVASALE	KWKTAIREA	QTFSR	MHVLLGMLDACIKWDMSAENARCKVCPKKGED 1200
Db	1141	KQWBEAKVASALE	KWKTAIREA	QTFSR	MHVLLGMLDACIKWDMSAENARCKVCPKKGED 1200
Qy	1201	DKLILCDECNKAF	HLFCLRP	ALYEP	DCWQCPACOPATARNRSGRNYTESASDSED 1260
Db	1201	DKLILCDECNKAF	HLFCLRP	ALYEP	DCWQCPACOPATARNRSGRNYTESASDSED 1260
Qy	1261	DESDEEEEEE	EEDEE	YEVAG	LRPRKTTIRGKHSVIPPAAARSRRPGKKPHSTRSQ 1320
Db	1261	DESDEEEEEE	EEDEE	YEVAG	LRPRKTTIRGKHSVIPPAAARSRRPGKKPHSTRSQ 1320
Qy	1321	PKAPPVDDAE	VDELVLQ	TKRSSR	RSQLELQKCEIILHKIVKYRFSWPPFPVTRDEAEDY 1380
Db	1321	PKAPPVDDAE	VDELVLQ	TKRSSR	RSQLELQKCEIILHKIVKYRFSWPPFPVTRDEAEDY 1380
Qy	1381	YDVIITHPMD	FOVQNK	CGSYRS	VOEFLTDMKQVFTNAEVNCRGSHVLS
Db	1381	YDVIITHPMD	FOVQNK	CGSYRS	VOEFLTDMKQVFTNAEVNCRGSHVLS
Qy	1441	VVLLHKHLPGH	PYVRK	KKPPDR	LAEDGDESEPAVGOSRDEDRRSREAEIQEWLQDTS 1500
Db	1441	VVLLHKHLPGH	PYVRK	KKPPDR	LAEDGDESEPAVGOSRDEDRRSREAEIQEWLQDTS 1500
Qy	1501	LYSAKINSK	DHNCFM	LVNTQ	FCMALTDVT 1531
Db	1501	LYSAKINSK	DHNCFM	LVNTQ	FCMALTDVT 1531
RESULT 3					
US-10-702-148-29					
; Sequence 29, Application US/10702148					
; Publication No. US20040063145A1					
; GENERAL INFORMATION:					
; APPLICANT: Jones, Michael H.					
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR					
; FILE REFERENCE: 06501-042002					
; CURRENT APPLICATION NUMBER: US/10/702,148					
; CURRENT FILING DATE: 2003-11-05					
; PRIOR APPLICATION NUMBER: PRIOR APPLICATION NUMBER: US/09/839,479					
; PRIOR FILING DATE: 2001-04-20					
; PRIOR APPLICATION NUMBER: US 09/418,710					
; PRIOR FILING DATE: 1999-10-15					
; PRIOR APPLICATION NUMBER: PCT/JP98/01783					
; PRIOR FILING DATE: 1998-04-17					
; PRIOR APPLICATION NUMBER: JP 9/310027					
; PRIOR FILING DATE: 1997-10-24					
; PRIOR APPLICATION NUMBER: JP 9/116570					
; PRIOR FILING DATE: 1997-04-18					
; NUMBER OF SEQ ID NOS: 72					
; SOFTWARE: FastSeq for Windows Version 4.0					
; SEQ ID NO 29					
; LENGTH: 1531					
; TYPE: PRT					
; ORGANISM: Homo sapiens					
US-10-702-148-29					
Query Match		100.0%; Score 7985; DB 15; Length 1531;			
Best Local Similarity		100.0%; Pred. No. 0;			
Matches 1531; Conservative		0; Mismatches 0; Indels 0; Gaps 0;			
Qy	1	MAPLGRKPPPLVNP	LPGEPPFTI	PHTQEA	PTREEYE
Db	1	MAPLGRKPPPLVNP	LPGEPPFTI	PHTQEA	PTREEYE
Qy	61	THKEAWEEQVVAELLKBEFFPAWYK	KL	VLEMHVHNTAS	LEKLVDTAWLEIMTKYAVGEEC 120

Db 61 THKEAWEEQVVAELLKEEFPAPWYEKLVLEMVHNTASLEKLVDTAWLEIMTKYAVGEEC 120
QY 121 DFEVGKEXMLVKYIKVIKIHLEKVDDEATEKSDGACDSSPKENSSQIAODHQKETT 180
Db 121 DFEVGKEXMLVKYIKVIKIHLEKVDDEATEKSDGACDSSPKENSSQIAODHQKETT 180
QY 181 KEDEGRRESINDRARRSPKLPSTLSKGERKWAPPKFLPHKYVDVKLQNEDKIISNVPADS 240
Db 181 KEDEGRRESINDRARRSPKLPSTLSKGERKWAPPKFLPHKYVDVKLQNEDKIISNVPADS 240
QY 241 LIRTERPPNKEIVRYFIRHNALRAGTGENAPWVEDELVKKYSLPKSFDFLLDPYKMT 300
Db 241 LIRTERPPNKEIVRYFIRHNALRAGTGENAPWVEDELVKKYSLPKSFDFLLDPYKMT 300
QY 301 LNPSTKRKTGSPDRPKSKSTDNSSLSPLNPKLWCHVHLKSLSGSPLKYNKSNK 360
Db 301 LNPSTKRKTGSPDRPKSKSTDNSSLSPLNPKLWCHVHLKSLSGSPLKYNKSNK 360
QY 361 SPEEHLEEMKMSPNKLTNPHIPKGPAPKPKGKSDPKLAKGRSGKILNGOKSTGN 420
Db 361 SPEEHLEEMKMSPNKLTNPHIPKGPAPKPKGKSDPKLAKGRSGKILNGOKSTGN 420
QY 421 SKSPKGLKTPKTKMQMTLLDMAKGTQMTAPRNSGTPRTSSPKHKHLPPLAALHLIA 480
Db 421 SKSPKGLKTPKTKMQMTLLDMAKGTQMTAPRNSGTPRTSSPKHKHLPPLAALHLIA 480
QY 481 YKENKREDKRSALSCVISTARLLSSDDRARLPEELRSLVQKRYELLEHKRWASWE 540
Db 481 YKENKREDKRSALSCVISTARLLSSDDRARLPEELRSLVQKRYELLEHKRWASWE 540
QY 541 EORKYLVKKREELKKLEKAKERREKEMLEKRYEDQELTGKMLPAFRVLDTPE 600
Db 541 EORKYLVKKREELKKLEKAKERREKEMLEKRYEDQELTGKMLPAFRVLDTPE 600
QY 601 GLPNTLFGDVAMVVEFLSCYSGLLLPDAQYPTAVSLMEALSADKGGFLYLNRLVILLQ 660
Db 601 GLPNTLFGDVAMVVEFLSCYSGLLLPDAQYPTAVSLMEALSADKGGFLYLNRLVILLQ 660
QY 661 TLLOTLQDEIADYDYGELGMKLSSEIPLTHSVSELVRLCLRRSDVOBESGSDTDNDKDS 720
Db 661 TLLOTLQDEIADYDYGELGMKLSSEIPLTHSVSELVRLCLRRSDVOBESGSDTDNDKDS 720
QY 721 AAFENEVDQDEPLEKLETSPELSEKQLQTLTALCHRLMTYSVDHMETRQMSAEL 780
Db 721 AAFENEVDQDEPLEKLETSPELSEKQLQTLTALCHRLMTYSVDHMETRQMSAEL 780
QY 781 WKERLAVLKEENDKKRAEKQKEMEAKNKENGKVENGLGKTDRKRIKVPQVDTEAE 840
Db 781 WKERLAVLKEENDKKRAEKQKEMEAKNKENGKVENGLGKTDRKRIKVPQVDTEAE 840
QY 841 DMSAVKSRRLIAIOAKKEREIOEREMKVKLROAEEERIIRKHAAAEKAFQEGIAKAL 900
Db 841 DMSAVKSRRLIAIOAKKEREIOEREMKVKLROAEEERIIRKHAAAEKAFQEGIAKAL 900
QY 901 VMRRTPIGTDNRHNYWLFSDVPGLFTTEKGWVHDSIDYRFNHCKDHTVSGDEDCPRS 960
Db 901 VMRRTPIGTDNRHNYWLFSDVPGLFTTEKGWVHDSIDYRFNHCKDHTVSGDEDCPRS 960
QY 961 KKANLGKNSMNTQHGTEAVETTTTPKQGNLWFLCDQSKELDLNLCLHPQGIRESQ 1020
Db 961 KKANLGKNSMNTQHGTEAVETTTTPKQGNLWFLCDQSKELDLNLCLHPQGIRESQ 1020
QY 1021 LKERLEKRYQDIISHIHLARKENLGLKSCDGNQELLNLFRLSDLIIEVATRLQXGGIYVEE 1080
Db 1021 LKERLEKRYQDIISHIHLARKENLGLKSCDGNQELLNLFRLSDLIIEVATRLQXGGIYVEE 1080
QY 1081 TSEFARVLSLEKLDQFGEVCVIALQASVTIKKFLQGFMAPKQRRKLQSEDSAKTEVEDEE 1140
Db 1081 TSEFARVLSLEKLDQFGEVCVIALQASVTIKKFLQGFMAPKQRRKLQSEDSAKTEVEDEE 1140
QY 1141 KQWVEAKVASALEKWKTAIREAQTFSRMHVLLGLMDACIKWDMSAENARCKVCPKKGED 1200
Db 1141 KQWVEAKVASALEKWKTAIREAQTFSRMHVLLGLMDACIKWDMSAENARCKVCPKKGED 1200

QY 1201 DKLILCDECNKAFHLCFLRPALYEVDPGECWCPACQATARNRSGRNVTEBSASEDS 1260
Db 1201 DKLILCDECNKAFHLCFLRPALYEVDPGECWCPACQATARNRSGRNVTEBSASEDS 1260
QY 1261 DESDEEEEBEEDYEVAGLRRLRPKTIIRGKHSVIPPAAARSGRRPGKKPHSTRSQ 1320
Db 1261 DESDEEEEBEEDYEVAGLRRLRPKTIIRGKHSVIPPAAARSGRRPGKKPHSTRSQ 1320
QY 1321 PKAPPVDDAEVDDELVLQTKRSSRQSLQKCEELHKKIVKYRFSWPFPVTRDEADY 1380
Db 1321 PKAPPVDDAEVDDELVLQTKRSSRQSLQKCEELHKKIVKYRFSWPFPVTRDEADY 1380
QY 1381 YDVLTHPMDFTQVQNKSCGYSYRVSQVEFLTDMKQVFTNAEVNCRGSHVLSMWKTEOCL 1440
Db 1381 YDVLTHPMDFTQVQNKSCGYSYRVSQVEFLTDMKQVFTNAEVNCRGSHVLSMWKTEOCL 1440
QY 1441 VVLLHKLHPGHPYVRRKKEKFPDLRAEDGDESEPAVQCSRDEDRRSREAEIQEWLQDTS 1500
Db 1441 VVLLHKLHPGHPYVRRKKEKFPDLRAEDGDESEPAVQCSRDEDRRSREAEIQEWLQDTS 1500
QY 1501 LYSAKINSKDHNCFMMLVNTQFCMALTDVT 1531
Db 1501 LYSAKINSKDHNCFMMLVNTQFCMALTDVT 1531

RESULT 4

US-09-839-479-27
; Sequence 27, Application US/09839479
; Publication No. US2002003979A1
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042002
; CURRENT APPLICATION NUMBER: US/09/839,479
; CURRENT FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JF98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 27
; LENGTH: 1527
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-839-479-27

Query Match 99.6%; Score 7955; DB 9; Length 1527;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 1527; Conservative 0; Mismatches 0; Indels 4; Gaps 1;

QY 1 MAPLLGRKPPFLVNPPLGCEEPFFIPIHTQEAFTREEYEARELRYSERIWTCKSGSSQL 60
Db 1 MAPLLGRKPPFLVNPPLGCEEPFFIPIHTQEAFTREEYEARELRYSERIWTCKSGSSQL 60
QY 61 THKEAWEEQVVAELLKEEFPAPWYEKLVLEMVHNTASLEKLVDTAWLEIMTKYAVGEEC 120
Db 61 THKEAWEEQVVAELLKEEFPAPWYEKLVLEMVHNTASLEKLVDTAWLEIMTKYAVGEEC 120
QY 121 DFEVGKEXMLVKYIKVIKIHLEKVDDEATEKSDGACDSSPKENSSQIAODHQKETT 180
Db 121 DFEVGKEXMLVKYIKVIKIHLEKVDDEATEKSDGACDSSPKENSSQIAODHQKETT 180
QY 181 KEDEGRRESINDRARRSPKLPSTLSKGERKWAPPKFLPHKYVDVKLQNEDKIISNVPADS 240
Db 181 KEDEGRRESINDRARRSPKLPSTLSKGERKWAPPKFLPHKYVDVKLQNEDKIISNVPADS 240
QY 241 LIRTERPPNKEIVRYFIRHNALRAGTGENAPWVEDELVKKYSLPKSFDFLLDPYKMT 300

Db 241 LIRTERPPNKEIVRYFIRHNALRAGTGENAPWVDELVKYSLSKFSDFLLDPYKMT 300
QY 301 LNPSTKRKNTGSPDRKPSKSKTDNSLSPLNPKLWCHVHLKXLSGSLPKVQNSKNSK 360
Db 301 LNPSTKRKNTGSPDRKPSKSKTDNSLSPLNPKLWCHVHLKXLSGSLPKVQNSKNSK 360
QY 361 SPEEHLSEEMKMSPNKLTNTHFI PKGPPAKKPGKHSKDKPLAKGRSKGIILNGOKSTGN 420
Db 361 SPEEHLSEEMKMSPNKLTNTHFI PKGPPAKKPGKHSKDKPLAKGRSKGIILNGOKSTGN 420
QY 421 SKSPKGLTPTKTKKQNTLLDMAGTKQKWTAPRNSGGTPTTSKPKHKLPPAALHLIA 480
Db 421 SKSPKGLTPTKTKKQNTLLDMAGTKQKWTAPRNSGGTPTTSKPKHKLPPAALHLIA 480
QY 481 YYKENKDRKDSALSVCISKTARLLSSDRARLPEELRSVLQRYELLEHKRWASMS 540
Db 481 YYKENKDRKDSALSVCISKTARLLSSDRARLPEELRSVLQRYELLEHKRWASMS 540
QY 541 EQRKEYLKKRBEELKKLKEKAKEREKEMLEKQRYEDQELTGKLPAPFLVDTP 600
Db 541 EQRKEYLKKRBEELKKLKEKAKEREKEMLEKQRYEDQELTGKLPAPFLVDTP 600
QY 601 GLPNTLFGDVAMVVEFLSCYSGILLPDAQYPTAVSLMEALSADKGGFLYLNRLVILQ 660
Db 601 GLPNTLFGDVAMVVEFLSCYSGILLPDAQYPTAVSLMEALSADKGGFLYLNRLVILQ 660
QY 661 TLLQTLQDEIAEDYGELGKMLSEIPLTLHVSSELVRLCLRSDDVOERSEGGSDTDNKS 720
Db 661 TLLQTLQDEIAEDYGELGKMLSEIPLTLHVSSELVRLCLRSDDVOERSEGGSDTDNKS 720
QY 721 AAFEDNEVDEFLKLETSSEFFELTSEELQILALTALCHRLMTYSQDHMETRQOMSABL 780
Db 721 AAFEDNEVDEFLKLETSSEFFELTSEELQILALTALCHRLMTYSQDHMETRQOMSABL 780
QY 777 WKERLAVLKEENDKRAEKQKREKEMAKNKGKVENGLKTKDRKRIVKFEPQVDTAE 840
Db 777 WKERLAVLKEENDKRAEKQKREKEMAKNKGKVENGLKTKDRKRIVKFEPQVDTAE 840
QY 841 DMISAVKSRLLAIQAKKEREIQEREMKVLBRQAEERIRKHAAAEKAFQEGIAKAL 900
Db 841 DMISAVKSRLLAIQAKKEREIQEREMKVLBRQAEERIRKHAAAEKAFQEGIAKAL 900
QY 901 VMRTPITGDRNHNRYLWFSDEVPLGFIKGGVHDSIDYRFNHCHKDHVTSDEDCPRS 960
Db 901 VMRTPITGDRNHNRYLWFSDEVPLGFIKGGVHDSIDYRFNHCHKDHVTSDEDCPRS 960
QY 961 KXANLGKNSMNTQHGTA TEVAVETTTPKQGNLWFLCDSQKXELDELLNCLHPOGIR 1020
Db 961 KXANLGKNSMNTQHGTA TEVAVETTTPKQGNLWFLCDSQKXELDELLNCLHPOGIR 1020
QY 1021 LKERLEKRYQDIHSHILARKENLGLKSCDGNQOELLNFLRSLIEVATRLQKGGLYVEE 1080
Db 1021 LKERLEKRYQDIHSHILARKENLGLKSCDGNQOELLNFLRSLIEVATRLQKGGLYVEE 1080
QY 1081 TSEFEARVTSLEKLDGFCVIALQASVTKKFLQGFMAPKQKRRKLQSDSDSAKTEBEVDE 1140
Db 1081 TSEFEARVTSLEKLDGFCVIALQASVTKKFLQGFMAPKQKRRKLQSDSDSAKTEBEVDE 1140
QY 1141 KKWBEAKVASALEKWKTAIRBAQTFSRMHVLLGLMDACIKWMSAENARCKVCPKKGED 1200
Db 1141 KKWBEAKVASALEKWKTAIRBAQTFSRMHVLLGLMDACIKWMSAENARCKVCPKKGED 1200
QY 1201 DKLLCDECNKAFHLCFLRPALEYVDGEMQCPACOPATARNRGRNVTESASEDSED 1260
Db 1201 DKLLCDECNKAFHLCFLRPALEYVDGEMQCPACOPATARNRGRNVTESASEDSED 1260
QY 1261 DESDEEBEEEEEEEDYEAVGLRLRPKRTIRGKHSVIPPAARSGRRPKPHSTRASQ 1320
Db 1261 DESDEEBEEEEEEEDYEAVGLRLRPKRTIRGKHSVIPPAARSGRRPKPHSTRASQ 1320
QY 1321 PKAPPVDDAEVDDELVTQTKRSRRSLELQKCEILHKIVKYRFSWPFREPVRTDAEDY 1380
Db 1321 PKAPPVDDAEVDDELVTQTKRSRRSLELQKCEILHKIVKYRFSWPFREPVRTDAEDY 1380

Db 1317 PKAPPVDDAEVDDELVTQTKRSRRSLELQKCEILHKIVKYRFSWPFREPVRTDAEDY 1376
QY 1381 YDVIITHPMDFOVQNKSCGYSYRSVQEFITDMKQVFTNAEVNCRGSHVLSQWVKTEOCL 1440
Db 1377 YDVIITHPMDFOVQNKSCGYSYRSVQEFITDMKQVFTNAEVNCRGSHVLSQWVKTEOCL 1436
QY 1441 VVLLHKHLPGHYPYRRKRRKKPDRLAEBEGDSEPEAVGOSRDEDRRSREAEIOEWLQDTS 1500
Db 1437 VVLLHKHLPGHYPYRRKRRKKPDRLAEBEGDSEPEAVGOSRDEDRRSREAEIOEWLQDTS 1496
QY 1501 LYSAKINSKDHNCFMVLVNTQFCMALTDVT 1531
Db 1497 LYSAKINSKDHNCFMVLVNTQFCMALTDVT 1527

RESULT 5

US-10-376-537-27
; Sequence 27, Application US/10376537
; Publication No. US20030224405A1
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042001
; CURRENT APPLICATION NUMBER: US/10/376,537
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US/09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 27
; LENGTH: 1527
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-376-537-27

Query Match 99.6%; Score 7955; DB 15; Length 1527;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 1527; Conservative 0; Mismatches 0; Indels 4; Gaps 1;
QY 1 MAPLGRKPPFLVNPPLGCEEPFTTIPHTQEAFTREEYEARLERYSERIWTCKSTGSSQL 60
Db 1 MAPLGRKPPFLVNPPLGCEEPFTTIPHTQEAFTREEYEARLERYSERIWTCKSTGSSQL 60
QY 61 THKEAWBEEQEAELKEEFPWYBKLVLWVHHNTASLEKLVDTAWLEIMTKYAVGESC 120
Db 61 THKEAWBEEQEAELKEEFPWYBKLVLWVHHNTASLEKLVDTAWLEIMTKYAVGESC 120
QY 121 DFEVGEKMLKVKIIVKIHPLKEDBEATEKKS DGACDPSDDKENSQIAQHQKKEVTV 180
Db 121 DFEVGEKMLKVKIIVKIHPLKEDBEATEKKS DGACDPSDDKENSQIAQHQKKEVTV 180
QY 181 KEDEGRRESINDRARRSPRKLPSTLSKGERKWPAPKFLPHKYVDVKLQNEDKIISNPADS 240
Db 181 KEDEGRRESINDRARRSPRKLPSTLSKGERKWPAPKFLPHKYVDVKLQNEDKIISNPADS 240
QY 241 LIRTERPNKEIVRYFIRHNALRAGTGENAPWVDELVKYSLSKFSDFLLDPYKMT 300
Db 241 LIRTERPNKEIVRYFIRHNALRAGTGENAPWVDELVKYSLSKFSDFLLDPYKMT 300
QY 301 LNPSTKRKNTGSPDRKPSKSKTDNSLSPLNPKLWCHVHLKXLSGSLPKVQNSKNSK 360
Db 301 LNPSTKRKNTGSPDRKPSKSKTDNSLSPLNPKLWCHVHLKXLSGSLPKVQNSKNSK 360
QY 361 SPEEHLSEEMKMSPNKLTNTHFI PKGPPAKKPGKHSKDKPLAKGRSKGIILNGOKSTGN 420
Db 361 SPEEHLSEEMKMSPNKLTNTHFI PKGPPAKKPGKHSKDKPLAKGRSKGIILNGOKSTGN 420

Qy 421 SKSPKGLKTKPTKMKQMTLLDMAKGTOKMTRAPNSGGTPTSSKPKHLLPPAALHLIA 480
Db |||||
Qy 421 SKSPKGLKTKPTKMKQMTLLDMAKGTOKMTRAPNSGGTPTSSKPKHLLPPAALHLIA 480
Db |||||
Qy 481 YYKENKDREKRSALSCVISTARTLLSSEDRARLPPEELRSVLQKRYELLEHKKRWASME 540
Db |||||
Qy 541 EQRKYLKKRBEELKKLKEKAKERREKEMLEERLEKQKRYEDQELTGKQLPAFRLVDTPE 600
Db |||||
Qy 541 EQRKYLKKRBEELKKLKEKAKERREKEMLEERLEKQKRYEDQELTGKQLPAFRLVDTPE 600
Db |||||
Qy 601 GLPNTLFGDVAMVFLSCYSGLLLPDAQYPTTAVSLMEALSADGGFLYLNRLVILLQ 660
Db |||||
Qy 601 GLPNTLFGDVAMVFLSCYSGLLLPDAQYPTTAVSLMEALSADGGFLYLNRLVILLQ 660
Db |||||
Qy 661 TLLQTLQDETAEDYGELGMLSEIPLTLHVSSELVRLCLRRSDVQSESESGSDTDNKNDS 720
Db -LLQTLQDETAEDYGELGMLSEIPLTLHVSSELVRLCLRRSDVQSESESGSDTDNKNDS 716
Qy 721 AAFENEVQDEPLEKLTSEFPPELTSEKLIQTLTALCHRLMTYVSQDHMETROQMSAEL 780
Db |||||
Qy 717 AAFENEVQDEPLEKLTSEFPPELTSEKLIQTLTALCHRLMTYVSQDHMETROQMSAEL 776
Db |||||
Qy 781 WKERLAVLKEENDKKRAEKQKBEKMEAKNKENGKVENGLGKTDKRRIVKPEPOVDTEAE 840
Db |||||
Qy 777 WKERLAVLKEENDKKRAEKQKBEKMEAKNKENGKVENGLGKTDKRRIVKPEPOVDTEAE 836
Db |||||
Qy 841 DMISAVKRRLLAIQAKKEREIQEREMKVKLEROAEEERIRKHAAAEKAFQEGIAKAKL 900
Db |||||
Qy 837 DMISAVKRRLLAIQAKKEREIQEREMKVKLEROAEEERIRKHAAAEKAFQEGIAKAKL 896
Db |||||
Qy 901 VNRPTPIGTDNRNNRYWLFSDVPGLFTEKGWVHDSIDYRNFHHCXDHVTSGDEBYCPRS 960
Db |||||
Qy 897 VNRPTPIGTDNRNNRYWLFSDVPGLFTEKGWVHDSIDYRNFHHCXDHVTSGDEBYCPRS 956
Db |||||
Qy 961 KKANLGKASMMTQHGTAETAVETTTPKQGNLWFLCDSQKELDELLNCLHPQGIRESQ 1020
Db |||||
Qy 957 KKANLGKASMMTQHGTAETAVETTTPKQGNLWFLCDSQKELDELLNCLHPQGIRESQ 1016
Db |||||
Qy 1021 LKERLEKRYQDIHSHILARKENLGLKSCDGNQELLNFLRSLDIEVATRLQKGGIYVEE 1080
Db |||||
Qy 1017 LKERLEKRYQDIHSHILARKENLGLKSCDGNQELLNFLRSLDIEVATRLQKGGIYVEE 1076
Db |||||
Qy 1081 TSEFARVLSLEKLDGFCVJALQASVIKKFLOGFMAPKQRRKLQSEDSAKTEBEVDEE 1140
Db |||||
Qy 1077 TSEFARVLSLEKLDGFCVJALQASVIKKFLOGFMAPKQRRKLQSEDSAKTEBEVDEE 1136
Db |||||
Qy 1141 KQWVEEAKVASALEKWKTAIREAQTFSRMHVLLGLMDACIKWDMSAENARCKVCPKGED 1200
Db |||||
Qy 1137 KQWVEEAKVASALEKWKTAIREAQTFSRMHVLLGLMDACIKWDMSAENARCKVCPKGED 1196
Db |||||
Qy 1201 DXLJILCDECNKAFHLCFLRPALYVDPGEWQCPACQPATARNRSGRNYTEESASEDSED 1260
Db |||||
Qy 1197 DXLJILCDECNKAFHLCFLRPALYVDPGEWQCPACQPATARNRSGRNYTEESASEDSED 1256
Db |||||
Qy 1261 DSDSEEBEEEEEBEEDYEVAGLRPRKTTIRGHSVIPPAARSGRRPGKPPHSTRRSQ 1320
Db |||||
Qy 1257 DSDSEEBEEEEEBEEDYEVAGLRPRKTTIRGHSVIPPAARSGRRPGKPPHSTRRSQ 1316
Db |||||
Qy 1321 PKAPVDDAEVDDELVLQTKRSSROSLELQKCEEILHKIKVYRFSWPPREPVTREAEADY 1380
Db |||||
Qy 1317 PKAPVDDAEVDDELVLQTKRSSROSLELQKCEEILHKIKVYRFSWPPREPVTREAEADY 1376
Db |||||
Qy 1381 YDVIITHPMDFQTVQNKSCSGSYRSVQEFITDMKQVFTNAEVYVNCRGSHVLSVMWTEOCL 1440
Db |||||
Qy 1377 YDVIITHPMDFQTVQNKSCSGSYRSVQEFITDMKQVFTNAEVYVNCRGSHVLSVMWTEOCL 1436
Db |||||
Qy 1441 VVLLHKLHPGHPYVRRKPKFPDRLAEDGDSPEAVGQSRDREDRRSREAEIQEWLQDTS 1500
Db |||||
Qy 1437 VVLLHKLHPGHPYVRRKPKFPDRLAEDGDSPEAVGQSRDREDRRSREAEIQEWLQDTS 1496
Db |||||
Qy 1501 LYSAKINSKDHNCFMVLVNTQFCMALTDVT 1531

Db 1497 LYSAKINSKDHNCFMVLVNTQFCMALTDVT 1527
|||
RESULT 6
US-10-702-148-27
; Sequence 27, Application US/10702148
; Publication No. US20040063145A1
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042002
; CURRENT APPLICATION NUMBER: US/10/702,148
; CURRENT FILING DATE: 2003-11-05
; PRIOR APPLICATION NUMBER: PRIOR APPLICATION NUMBER: US/09/839,479
; PRIOR FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 27
; LENGTH: 1527
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-702-148-27

Query Match 99.6%; Score 7955; DB 15; Length 1527;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 1527; Conservative 0; Mismatches 0; Indels 4; Gaps 1;

Qy 1 MAPLLGRKPPFLVNPPLGCEBFFFTPIPTQEAFTREEYEARLERYSERIWTCKSTGSSQL 60
Db 1 MAPLLGRKPPFLVNPPLGCEBFFFTPIPTQEAFTREEYEARLERYSERIWTCKSTGSSQL 60
Qy 61 THKEAWEEQEAELLKEEFPAYEKLVLVEMVHNTASLEKLVDTAMLEIMTKYAVGSEC 120
Db 61 THKEAWEEQEAELLKEEFPAYEKLVLVEMVHNTASLEKLVDTAMLEIMTKYAVGSEC 120
Qy 121 DFEVGEKMLVKVIVKHPLEKVDDEATEKSDGACDSPSSDKENSSQIAODHOKETVV 180
Db 121 DFEVGEKMLVKVIVKHPLEKVDDEATEKSDGACDSPSSDKENSSQIAODHOKETVV 180
Qy 181 KEDEGRRESINDRARRSRPKLPTSLKKGKRWAPPKFLPHKYDVKLQNEDKIISNVPADS 240
Db 181 KEDEGRRESINDRARRSRPKLPTSLKKGKRWAPPKFLPHKYDVKLQNEDKIISNVPADS 240
Qy 241 LIRTERPPNKEIVRYFIRHNALRAGTGENAPWVVEDELVKKYSLPSPKSFDFLLDPKYMT 300
Db 241 LIRTERPPNKEIVRYFIRHNALRAGTGENAPWVVEDELVKKYSLPSPKSFDFLLDPKYMT 300
Qy 301 LNPSTKRKNTGSPDRKPSKSKTDNSSLSPNPKLWCHVHLKKSLSGSLPKVKNKSNKSK 360
Db 301 LNPSTKRKNTGSPDRKPSKSKTDNSSLSPNPKLWCHVHLKKSLSGSLPKVKNKSNKSK 360
Qy 361 SPEEHLBEMMKWMSPNKLTHTNFHIPKKGPPAKKPKGKSHDKPLKAKGRSGKILNGOKSTGN 420
Db 361 SPEEHLBEMMKWMSPNKLTHTNFHIPKKGPPAKKPKGKSHDKPLKAKGRSGKILNGOKSTGN 420
Qy 421 SKSPKGLKTKPTKMKQMTLLDMAKGTOKMTRAPNSGGTPTSSKPKHLLPPAALHLIA 480
Db 421 SKSPKGLKTKPTKMKQMTLLDMAKGTOKMTRAPNSGGTPTSSKPKHLLPPAALHLIA 480
Qy 481 YYKENKDREKRSALSCVISTARTLLSSEDRARLPPEELRSVLQKRYELLEHKKRWASME 540
Db 481 YYKENKDREKRSALSCVISTARTLLSSEDRARLPPEELRSVLQKRYELLEHKKRWASME 540
Qy 541 EQRKYLKKRBEELKKLKEKAKERREKEMLEERLEKQKRYEDQELTGKQLPAFRLVDTPE 600

Db 541 EORKEYLKKRBEELKKLKEKAKERREKEMLEKQKRYEDELTKGNLPAFLVDTP 600
Qy 601 GLPNTLFGDVAMVVEFLSCYSGILLPDAQYPTITAVSLMEALSADKGGFLYLNRLVILQ 660
Db 601 GLPNTLFGDVAMVVEFLSCYSGILLPDAQYPTITAVSLMEALSADKGGFLYLNRLVI 657
Qy 661 TLLOTLQDEIAEDYGEIEMKLSERPLTLHSHSELVRLCLRSDSDVQEESEGSDTDNKS 720
Db 658 -LLQTLQDEIAEDYGEIEMKLSERPLTLHSHSELVRLCLRSDSDVQEESEGSDTDNKS 716
Qy 721 AAFEDNEVDEFLKLETFEPELSEELKQILTLCHILMTYSVOOHMETROQMSAEL 780
Db 717 AAFEDNEVDEFLKLETFEPELSEELKQILTLCHILMTYSVOOHMETROQMSAEL 776
Qy 781 WKERLAVLKEENDKKRAEKQKKEKEMAKNKENGKVENGLKTDKRRKIVKFPQVDTEAE 840
Db 777 WKERLAVLKEENDKKRAEKQKKEKEMAKNKENGKVENGLKTDKRRKIVKFPQVDTEAE 836
Qy 841 DMISAVKSRLLAIQAKKREIIOEREMKVKLRQRAEERIRKHAAAEEKAFQEGIAKAL 900
Db 837 DMISAVKSRLLAIQAKKREIIOEREMKVKLRQRAEERIRKHAAAEEKAFQEGIAKAL 896
Qy 901 VMRTPIGTDNRHNRWYLFSDVPGLFTEKGWVHDSIDYRFNHCKDHTVSGDEYCPRS 960
Db 897 VMRTPIGTDNRHNRWYLFSDVPGLFTEKGWVHDSIDYRFNHCKDHTVSGDEYCPRS 956
Qy 961 KXANLGKXANMTQGTATEVAVETTPKQGNLWFLCDSQKELDELLNCLHPQGIRESQ 1020
Db 957 KXANLGKXANMTQGTATEVAVETTPKQGNLWFLCDSQKELDELLNCLHPQGIRESQ 1016
Qy 1021 LKERLEKRYQDIHSHILARKPNLGLKSCDGNQELLNFLRSLIEVATRLQKGLGYVEE 1080
Db 1017 LKERLEKRYQDIHSHILARKPNLGLKSCDGNQELLNFLRSLIEVATRLQKGLGYVEE 1076
Qy 1081 TSEFARVLSLEKLDGFCVITALQASVIKKFLOGFMAPKQKRLQSEDSAKTEBVD 1140
Db 1077 TSEFARVLSLEKLDGFCVITALQASVIKKFLOGFMAPKQKRLQSEDSAKTEBVD 1136
Qy 1141 KQWVEAKVASALEKWKTAIRAQTFSRMHVLLGMLDACIKWDMSAENARCKVCPKGED 1200
Db 1137 KQWVEAKVASALEKWKTAIRAQTFSRMHVLLGMLDACIKWDMSAENARCKVCPKGED 1196
Qy 1201 DKLLCDECNKAFHFLCLRALPYEVPDGEWQCPACQATARNRSGRNYTEESASEDSED 1260
Db 1197 DKLLCDECNKAFHFLCLRALPYEVPDGEWQCPACQATARNRSGRNYTEESASEDSED 1256
Qy 1261 DESDEEEEEEEDYEVAGLRLPRKTTIRGHSHVIPPAAARGRRPGKPKPHSTRRSQ 1320
Db 1257 DESDEEEEEEEDYEVAGLRLPRKTTIRGHSHVIPPAAARGRRPGKPKPHSTRRSQ 1316
Qy 1321 KPAPPVDDAEVDLQTKRSRRSLOLEKCEEILHKIKVYRFSWPPREPVTREAEADY 1380
Db 1317 KPAPPVDDAEVDLQTKRSRRSLOLEKCEEILHKIKVYRFSWPPREPVTREAEADY 1376
Qy 1381 YDVITHPMDFTQVQNCSCGYSRYSQEFITDMKQVFTNAEVYVNCGRSHVLSCMVQTEOCL 1440
Db 1377 YDVITHPMDFTQVQNCSCGYSRYSQEFITDMKQVFTNAEVYVNCGRSHVLSCMVQTEOCL 1436
Qy 1441 VYLLHKHLPHPYVRKPKKFPDRLAEDGDEPEAVGOSRDEDRRSREAEIOEWLQDTS 1500
Db 1437 VYLLHKHLPHPYVRKPKKFPDRLAEDGDEPEAVGOSRDEDRRSREAEIOEWLQDTS 1496
Qy 1501 LYSAKINSKDHNCFMMLVNTQFCMALTDFTVT 1531
Db 1497 LYSAKINSKDHNCFMMLVNTQFCMALTDFTVT 1527

RESULT 7

US-09-839-479-68
; Sequence 68, Application US/09839479
; Publication No. US20020039779A1
; GENERAL INFORMATION:

; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042002
; CURRENT APPLICATION NUMBER: US/09/839,479
; CURRENT FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 68
; LENGTH: 1525
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-839-479-68

Query Match 98.6%; Score 7876; DB 9; Length 1525;
Best Local Similarity 99.3%; Pred. No. 0;
Matches 1521; Conservative 1; Mismatches 2; Indels 8; Gaps 5;
Qy 1 MAPLGRKPPFLVNPPLGEEPPFTTIPHTQEAFTREEYEARLERYSERIWTCKTGSSQL 60
Db 1 MAPLGRKPPFLVNPPLGEEPPFTTIPHTQEAFTREEYEARLERYSERIWTCKTGSSQL 60
Qy 61 THKEAWBEEQVAAELKEEFPNAYEKLVMVHHNTASLEKLVDTAWLIMTKYAVGEEC 120
Db 61 THKEAWBEEQVAAELKEEFPNAYEKLVMVHHNTASLEKLVDTAWLIMTKYAVGEEC 120
Qy 121 DFEVCKEKLKVKIHKIHPLEKVDSEATEKSDGACDSSDKENSSQIAQDHQKKTVV 180
Db 121 DFEVCKEKLKVKIHKIHPLEKVDSEATEKSDGACDSSDKENSSQIAQDHQKKTVV 180
Qy 181 KEDEGRRESINDRARRSPKLPSTSLKGERKWAPPKFLPHKYDVKLQNEDEKIIISNVPADS 240
Db 181 KEDEGRRESINDRARRSPKLPSTSLKGERKWAPPKFLPHKYDVKLQNEDEKIIISNVPADS 240
Qy 241 LIRTERPNKEIVRYFIRHNALRAGTGNAPWVDELVKYSLSKPSKFSDFLLDPKYQWT 300
Db 241 LIRTERPNKEIVRYFIRHNALRAGTGNAPWVDELVKYSLSKPSKFSDFLLDPKYQWT 300
Qy 301 LNPSTRKNTGSPDRKPSKSTONSLSPLNPKLACHVHLKXSLSGSPLKVKXSKNSK 360
Db 301 LNPSTRKNTGSPDRKPSKSTONSLSPLNPKLACHVHLKXSLSGSPLKVKXSKNSK 360
Qy 361 SPEEHLEEMKXMSPNKLTNPHIPKGGPPAKKPGKHSKDKPLKAKGRSKGILNGOKSTGN 420
Db 361 SPEEHLEEMKXMSPNKLTNPHIPKGGPPAKKPGKHSKDKPLKAKGRSKGILNGOKSTGN 420
Qy 421 SKSPKKGKLTPTKTKQMTLLDMAKGTQMTAPRNSGGTPTTSKPKHKLPPAALHLTA 480
Db 421 SKSPKKGKLTPTKTKQMTLLDMAKGTQMTAPRNSGGTPTTSKPKHKLPPAALHLTA 480
Qy 481 YKKNKORDEKRSALSCVISTARLLSSEDRARLPEELRSLOKRYELLEHKKRWASME 540
Db 481 YKKNKORDEKRSALSCVISTARLLSSEDRARLPEELRSLOKRYELLEHKKRWASME 540
Qy 541 EORKEYLKKRBEELKKLKEKAKERREKEMLEKQKRYEDELTKGNLPAFLVDTP 600
Db 541 EORKEYLKKRBEELKKLKEKAKERREKEMLEKQKRYEDELTKGNLPAFLVDTP 600
Qy 601 GLPNTLFGDVAMVVEFLSCYSGILLPDAQYPTITAVSLMEALSADKGGFLYLNRLVILQ 660
Db 601 GLPNTLFGDVAMVVEFLSCYSGILLPDAQYPTITAVSLMEALSADKGGFLYLNRLVI 656
Qy 661 TLLOTLQDEIAEDYGEIEMKLSERPLTLHSHSELVRLCLRSDSDVQEESEGSDTDNKS 720
Db 657 -LLQTLQDEIAEDYGEIEMKLSERPLTLHSHSELVRLCLRSDSDVQEESEGSDTDNKS 715

Qy	721	AAFPEDNEVQDEFFLEKLETSBFFELTSEBKLIQILTALCHRILMTYSVODHMETRQMSAEL	780
Db	716	AAFPEDNEVQDEFFLEKLETSBFFELTSEBKLIQILTALCHRILMTYSVODHMETRQMSAEL	775
Qy	781	WKERLAVLKEENDKKRAEKOKRMEAKNKXNGKVENGLKTKRKRIVKFPEQVDTAE	840
Db	776	WKERLAVLKEENDKKRAEKOKRMEAKNKXNGKVENGLKTKRKRIVKFPEQVDTAE	835
Qy	841	DMISAVKSRLLAIQAKKEREIOEREMKVKLERQABEERIRKHKAAAEKAFQSGIAKAKL	900
Db	836	DMISAVKSRLLAIQAKKEREIOEREMKVKLERQABEERIRKHKAAAEKAFQSGIAKAKL	895
Qy	901	VMRPTPTGDRNHNRYWLFSDVPGLFIEKGWHDSIDYRPNHCKDHTVSGDEDYCPRS	960
Db	896	VMRPTPTGDRNHNRYWLFSDVPGLFIEKGWHDSIDYRPNHCKDHTVSGDEDYCPRS	955
Qy	961	KKANLGKVASNTOHGATATAVAVETTPKOGQNLWFLCDSQKELDELLNCLHPQGIRRESQ	1020
Db	956	KKANLGKVASNTOHGATATAVAVETTPKOGQNLWFLCDSQKELDELLNCLHPQGIRRESQ	1015
Qy	1021	LKERLEKRYQDIITHSLHARKPNLGLKSCDGNQBELNFLRSDLIEVATRILOKGLGLGYEE	1080
Db	1016	LKERLEKRYQDIITHSLHARKPNLGLKSCDGNQBELNFLRSDLIEVATRILOKGLGLGYEE	1075
Qy	1081	TSEFEARVILEKLUKDFGECVIALOASVILEKFLQGFMAPOKRRKLOSEDSAKTEEVDEE	1140
Db	1076	TSEFEARVILEKLUKDFGECVIALOASVILEKFLQGFMAPOKRRKLOSEDSAKTEEVDEE	1135
Qy	1141	KQWVEEAKVASALEKWKTAIREAQTFSRMHVLLGMLDACIKWDMSAENARCKVCPKKGED	1200
Db	1136	KQWVEEAKVASALEKWKTAIREAQTFSRMHVLLGMLDACIKWDMSAENARCKVCPKKGED	1195
Qy	1201	DKLILCDECNKAHFLCRLPALYEVPOGEWQCPACQPATARRNSRGNHYTESASEDSE	1260
Db	1196	DKLILCDECNKAHFLCRLPALYEVPOGEWQCPACQPATARRNSRGNHYTESASEDSE	1254
Qy	1261	DESDEEEEEEEEEEDYEVAGURLRPRTIRGKHSVIPPAARSGRRPCKKPHSTERSQ	1320
Db	1255	DESDEEEEEEEEEEDYEVAGURLRPRTIRGKHSVIPPAARSGRRPCKKPHSTERSQ	1314
Qy	1321	PKAPPVDDAEVDELVLQTKGSSRRQSLELOKCEBILHKIVKYRFSWPPREPVRTRDEAEDY	1380
Db	1315	PKAPPV-DAEVDELVLQTKGSSRRQSLELOKCEBILHKIVKYRFSWPPREPVRTRDEAEDY	1373
Qy	1381	YDVITHPMDQTVQNKSCSGSYRSVQFBLTDMQKQVFNAEYVNCGRSHVLSCHWKTQOCL	1440
Db	1374	YDVITHPMDQTVQNKSCSGSYRSVQFBLTDMQKQVFNAEYVNCGRSHVLSCHWKTQOCL	1433
Qy	1441	VVLHKLHPGHPYVRRKXKFPDLAEDGDSPEAVQSGRDEDRRRREAEIOEWLQDTS	1500
Db	1434	VVLHKLHPGHPYVRRKXKFPDLAEDGDSPEAVQSGRDEDRRRREAEIOEWLQDTS	1493
Qy	1501	LY-SAKINSKDHNCFMMLVNTQFCWALTDVT	1531
Db	1494	LYASAKINSKDHNCFMMLVNTQFCWALTDVT	1525

RESULT 8

	KZSDI.9	
	US-10-376-537-69	
	: Sequence 69, Application US/10376537	
	: Publication No. US20030224405A1	
	: GENERAL INFORMATION:	
	: APPLICANT: Jones, Michael H.	
	: TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR	
	: FILE REFERENCE: 06501-042001	
	: CURRENT APPLICATION NUMBER: US/10/376,537	
	: CURRENT FILING DATE: 2003-02-28	
	: PRIOR APPLICATION NUMBER: US/09/418,710	
	: PRIOR FILING DATE: 1999-10-15	
	: PRIOR APPLICATION NUMBER: PCT/JF98/01783	
	: PRIOR FILING DATE: 1998-04-17	
	: PRIOR APPLICATION NUMBER: JP 9/310027	
	: PRIOR FILING DATE: 1997-10-24	

661	TLLOTTLLQDEIAEDYGELGNKMLSEIPLTLHVSSELVRLCLRRSDVOEESGSDTDDNKDS	Qy	720
D6	-LLOTTLLQDEIAEDYGELGNKMLSEIPLTLHVSSELVRLCLRRSDVOEESGSDTDDNKDS	D6	715
721	AAPEDNEVODEPLEKETSETSEFFELTSEEKLIILITLACHRILMTYSVQDHWMETRQQMSAEL	Qy	780
D6	AAPEDNEVODEPLEKETSETSEFFELTSEEKLIILITLACHRILMTYSVQDHWMETRQQMSAEL	D6	775
781	WKERLAVLKENDKKRAEKQKRKMEAKNKENGVENGLGKTDRKKRI VKFPQVDTEAE	Qy	840
D6	WKERLAVLKENDKKRAEKQKRKMEAKNKENGVENGLGKTDRKKRI VKFPQVDTEAE	D6	835
841	DMISAVKSRRLLAIQAKKEREIOEREMKVKLEROAEEERIIRKHAAAEEKAFOGGIAKAL	Qy	900
D6	DMISAVKSRRLLAIQAKKEREIOEREMKVKLEROAEEERIIRKHAAAEEKAFOGGIAKAL	D6	895

Qy	901	VMRTPITGTDNRHNRYYLWFLSDEVPGLFTIEKGWVHDSIDYRFNHHCKDHTVSGDBDYCPRS	960
Db	896	VMRTPITGTDNRHNRYYLWFLSDEVPGLFTIEKGWVHDSIDYRFNHHCKDHTVSGDBDYCPRS	955
Qy	961	KKANLGKGNASMTQHGTA TEAVETTPKQGNLWFLCDSQKEJDELLNCLHPGIGRESQ	1020
Db	956	KKANLGKGNASMTQHGTA TEAVETTPKQGNLWFLCDSQKEJDELLNCLHPGIGRESQ	1015
Qy	1021	LKERLEKRYQDI IHSIHLARKPNLGLKSCDGNQBELNFLRSDLIEVATRLQKGGGLGYVEE	1080
Db	1016	LKERLEKRYQDI IHSIHLARKPNLGLKSCDGNQBELNFLRSDLIEVATRLQKGGGLGYVEE	1075
Qy	1081	TSEFEARVISLEKLDKDFGECVIALQASVKKPLOGFWAPKQKRKLQSEDSAKTEEVDEE	1140
Db	1076	TSEFEARVISLEKLDKDFGECVIALQASVKKPLOGFWAPKQKRKLQSEDSAKTEEVDEE	1135
Qy	1141	KMWVEEAKVASALEKWKTAIREAOTFSRMHVLLGMLDACIKWDMSAENARCKVCPKXGED	1200
Db	1136	KMWVEEAKVASALEKWKTAIREAOTFSRMHVLLGMLDACIKWDMSAENARCKVCPKXGED	1195
Qy	1201	DKLILCDECNKAFHLCFLRPALYEVDPGEWCPCQCAPATARNRSGRNYTEESASEDSE	1260
Db	1196	DKLILCDECNKAFHLCFLRPALYEVDPGEWCPCQCAPATARNRSGRNYTEESASEDSE	1254
Qy	1261	DESEDEEEEEEBEEDYEVAGILRLRPKTRIRGKHSVIPPAARSGRRPGKKPHSTRRSQ	1320
Db	1255	DESEDEEEEEEBEEDYEVAGILRLRPKTRIRGKHSVIPPAARSGRRPGKKPHSTRRSQ	1314
Qy	1321	PKAPPVDDAEVDLVLQTKRSRRQSLELOKCEBILHKIVKYRFSWFPFPVTRDEAEDY	1380
Db	1315	PKAPPV-DAEVDLVLQTKRSRRQSLELOKCEBILHKIVKYRFSWFPFPVTRDEAEDY	1373
Qy	1381	YDVITHPMDFTQVQKSCGSYRSVQBFLLTDMKQVFTNAEYVNCRGSHVLSCMVKTBOCL	1440
Db	1374	YDVITHPMDFTQVQKSCGSYRSVQBFLLTDMKQVFTNAEYVNCRGSHVLSCMVKTBOCL	1433
Qy	1441	VVLHKLHLPCHPYRRKKKPPDLAEDGDSPEAVGQSRDEDRRSREAEIOEWLQDTS	1500
Db	1434	VVLHKLHLPCHPYRRKKKPPDLAEDGDSPEAVGQSRDEDRRSREAEIOEWLQDTS	1493
Qy	1501	LY-SAKINSKDHNCFMMLVNTQFCWALTDVT	1531
Db	1494	LYASAKINSKDHNCFMMLVNTQFCWALTDVT	1525
RESULT 9			
US-10-702-148-68			
; Sequence 68, Application US/10702148			
; Publication No.: US20040063145A1			
; GENERAL INFORMATION:			
; APPLICANT: Jones, Michael H.			
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR			
; FILE REFERENCE: 06501-042002			
; CURRENT APPLICATION NUMBER: US/10/702,148			
; CURRENT FILING DATE: 2003-11-05			
; PRIOR APPLICATION NUMBER: PRIOR APPLICATION NUMBER: US/09/839,479			
; PRIOR FILING DATE: 2001-04-20			
; PRIOR APPLICATION NUMBER: US 09/418,710			
; PRIOR FILING DATE: 1999-10-15			
; PRIOR APPLICATION NUMBER: PCT/JP98/01783			
; PRIOR FILING DATE: 1998-04-17			
; PRIOR APPLICATION NUMBER: JP 9/310027			
; PRIOR FILING DATE: 1997-10-24			
; PRIOR APPLICATION NUMBER: JP 9/116570			
; PRIOR FILING DATE: 1997-04-18			
; NUMBER OF SEQ ID NOS: 72			
; SOFTWARE: FastSeq for Windows Version 4.0			
; SEQ ID NO 68			
; LENGTH: 1525			
; TYPE: PRT			
; ORGANISM: Homo sapiens			
US-10-702-148-68			

Query Match	98.6%	Score 7876;	DB 15;	Length 1525;
Best Local Similarity	99.3%	Pred. No. 0;		
Matches 1521;	Conservative 1;	Mismatches 2;	Indels 8;	Gaps 5;
Qy	1	MAPLLGRKPPLVNPPLPGSEPPFTTIPHQTEAFTRBEYEARELRYSERIWTCKSTGSSQL	60	
Db	1	MAPLLGRKPPLVNPPLPGSEPPFTTIPHQTEAFTRBEYEARELRYSERIWTCKSTGSSQL	60	
Qy	61	THKEAWEEEOEVAELLKEEFPANYPEKLVLMEVHHNTASLEKLVDTAWLEIWTMYAVGEEC	120	
Db	61	THKEAWEEEOEVAELLKEEFPANYPEKLVLMEVHHNTASLEKLVDTAWLEIWTMYAVGEEC	120	
Qy	121	DFEVGKEMKLVKVIKIHPLKVDSEATEKKSDGACDSPSSDKENSQIAQDHQKKEVTV	180	
Db	121	DFEVGKEMKLVKVIKIHPLKVDSEATEKKSDGACDSPSSDKENSQIAQDHQKKEVTV	180	
Qy	181	KEDEGRRESINDRARRSPKLPSTLKKGERKWAAPPFLPHKYDVKLQNEBKII SNVPADS	240	
Db	181	KEDEGRRESINDRARRSPKLPSTLKKGERKWAAPPFLPHKYDVKLQNEBKII SNVPADS	240	
Qy	241	LIRTERPPNKEI VRYFIRHNALRAAGTGENAPWVVEDELVKYSLPSKFSDFLLDPYKMT	300	
Db	241	LIRTERPPNKEI VRYFIRHNALRAAGTGENAPWVVEDELVKYSLPSKFSDFLLDPYKMT	300	
Qy	301	LNPSSTRKNTGSPDRKPSKSKTNDNSLSPLNPKLWCHVHLKKSISGSPLVKNSKNKSK	360	
Db	301	LNPSSTRKNTGSPDRKPSKSKTNDNSLSPLNPKLWCHVHLKKSISGSPLVKNSKNKSK	360	
Qy	361	SPEHLEBEMKMWSPNKLHNTNPHI PKKGGPPAKPGHSDKPLAKAGRSKGI LNKQKSTGN	420	
Db	361	SPEHLEBEMKMWSPNKLHNTNPHI PKKGGPPAKPGHSDKPLAKAGRSKGI LNKQKSTGN	420	
Qy	421	SKSPKXGLKTPKTKMQMTLLDMWAKGTQKWTRAPRNSGGTPRTSSKPHKHLPPAALHLIA	480	
Db	421	SKSPKXGLKTPKTKMQMTLLDMWAKGTQKWTRAPRNSGGTPRTSSKPHKHLPPAALHLIA	480	
Qy	481	YYKENKODREKRSALSCVSKTARLLSSEDRARLPPEELRSLVOKRYELLEHKKRWASME	540	
Db	481	YYKENKODREKRSALSCVSKTARLLSSEDRARLPPEELRSLVOKRYELLEHKKRWASME	540	
Qy	541	EQRKEYLKKRBEELKKLKEKAKEREKEMLERLEKQRYEDELGTGKNLPAPFLVDTP	600	
Db	541	EQRKEYLKKRBEELKKLKEKAKEREKEMLERLEKQRYEDELGTGKNLPAPFLVDTP	600	
Qy	601	GLPNTLFGDVAMVVEFLSCYSGLLLPDAQYPI TAVSIMEALSADKGGFLYLNRLVILLQ	660	
Db	601	GLPNTLFGDVAMVVEFLSCYSGLLLPDAQYPI TAVSIMEAL-ADKGGFLYLNRLVI---	656	
Qy	661	TLIQTLLQDEIAEDYGEIGMKLSEIPLTLHSUSELYRLCLRRSDVOEESGSDTDNKKOS	720	
Db	657	-LLIQTLLQDEIAEDYGEIGMKJSEIPLTLHSUSELYRLCLRRSDVOEESGSDTDNKKOS	715	
Qy	721	AAEPEDNEVQDEFLKLETSSEFFELTSEEKLIQILTALCHRLIMTYSVDHMETRQOMSAEL	780	
Db	716	AAEPEDNEVQDEFLKLETSSEFFELTSEEKLIQILTALCHRLIMTYSVDHMETRQOMSAEL	775	
Qy	781	WKERLAVLKEENDKQAEKQKREKEMEAQKNKENGKVENGLGKTDKRRIVKVFEPQVDEAE	840	
Db	776	WKERLAVLKEENDKQAEKQKREKEMEAQKNKENGKVENGLGKTDKRRIVKVFEPQVDEAE	835	
Qy	841	DMISAVKSRLLAIQAKKEREIOEREMKVKLEPQAEERIRKHAAAEKAFQGGIAKAKL	900	
Db	836	DMISAVKSRLLAIQAKKEREIOEREMKVKLEPQAEERIRKHAAAEKAFQGGIAKAKL	895	
Qy	901	VMRTPITGTDNRHNRYYLWFLSDEVPGLFIEKGWVHDSIDYRFNHHCKDHTVSGDEDCPRS	960	
Db	896	VMRTPITGTDNRHNRYYLWFLSDEVPGLFIEKGWVHDSIDYRFNHHCKDHTVSGDEDCPRS	955	
Qy	961	KKANLGKGNASMTQHGTA TEAVETTPKQGNLWFLCDSQKEJDELLNCLHPGIGRESQ	1020	
Db	956	KKANLGKGNASMTQHGTA TEAVETTPKQGNLWFLCDSQKEJDELLNCLHPGIGRESQ	1015	

RESULT 9

US-10-702-148-68	Db	601	GLPNTLFGDVAMVVEFLSCYSGLLLPDAQYPIITAVSLMEAL-ADKGGFLYNRVLYI---	656
Sequence 68, Application US/10702148				
Publication No. US20040063145A1				
GENERAL INFORMATION:				
APPLICANT: Jones, Michael H.				
TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR				
FILE REFERENCE: 06501-042002				
CURRENT APPLICATION NUMBER: US/10/702,148				
CURRENT FILING DATE: 2003-11-05				
PRIOR APPLICATION NUMBER: PRIOR APPLICATION NUMBER: US/09/839,479				
PRIOR FILING DATE: 2001-04-20				
PRIOR APPLICATION NUMBER: US 09/418,710				
PRIOR FILING DATE: 1999-10-15				
PRIOR APPLICATION NUMBER: PCT/JP98/01783				
PRIOR FILING DATE: 1998-04-17				
PRIOR APPLICATION NUMBER: JP 9/310027				
PRIOR FILING DATE: 1997-10-24				
PRIOR APPLICATION NUMBER: JP 9/116570				
PRIOR FILING DATE: 1997-04-18				
NUMBER OF SEQ ID NOS: 72				
SOFTWARE: FastSeq for Windows Version 4.0				
SEQ ID NO 68				
LENGTH: 1525				
TYPE: Prt				
ORGANISM: Homo sapiens				
US-10-702-148-68				

QY 1021 LKERLEKRYQDIIHSHLARKENLGLKSCDGNQOELLNFRSLDIEVATRLQKGLGYVEE 1080
Db 1016 LKERLEKRYQDIIHSHLARKENLGLKSCDGNQOELLNFRSLDIEVATRLQKGLGYVEE 1075
QY 1081 TSEFEARVLSLEKLDGECVIALQASVIKKFLOGFMAPKQKRRKLQSEDSAKTEEVDEE 1140
Db 1076 TSEFEARVLSLEKLDGECVIALQASVIKKFLOGFMAPKQKRRKLQSEDSAKTEEVDEE 1135
QY 1141 KQWBEAKVASALEKWKTAIRAQAQFTRSMHVLGMLDACIKWDMSAENARCKVCPKKGED 1200
Db 1136 KQWBEAKVASALEKWKTAIRAQAQFTRSMHVLGMLDACIKWDMSAENARCKVCPKKGED 1195
QY 1201 DKLLCDECNKAFHLCFLRPALYEPDGEWQCPACOPATARNRSGRNYTEESASEDS 1260
Db 1196 DKLLCDECNKAFHLCFLRPALYEPDGEWQCPACOPATARNRSGRNYTEESASEDS 1254
QY 1261 DESDEEEEEEEDDEEYEVAGLRPRKTIKRGKHSVIPPAAARGRRPGKKPHSTRSQ 1320
Db 1255 DESDEEEEEEEDDEEYEVAGLRPRKTIKRGKHSVIPPAAARGRRPGKKPHSTRSQ 1314
QY 1321 KPAPVDDAEVDLVLQTKRSGRRSLELQKCEILHKIKVYRFSWPFPVTRDEADY 1380
Db 1315 KPAPV-DAEVDLVLQTKRSGRRSLELQKCEILHKIKVYRFSWPFPVTRDEADY 1373
QY 1381 YDVITHPMDQTVQNKSCGYSRYSQEFITDMKQVFTNAEYVNCRGSHVLSQWVTEOCL 1440
Db 1374 YDVITHPMDQTVQNKSCGYSRYSQEFITDMKQVFTNAEYVNCRGSHVLSQWVTEOCL 1433
QY 1441 VLLHKLHPGHPYVRKPKKFPDRDLAEDGSEPEAVGQSDRDRRREAEIQEWLQDTS 1500
Db 1434 VLLHKLHPGHPYVRKPKKFPDRDLAEDGSEPEAVGQSDRDRRREAEIQEWLQDTS 1493
QY 1501 LY-SAKINSKOHNCFMFLVNTQFCMALTDVT 1531
Db 1494 LYASAKINSKOHNCFMFLVNTQFCMALTDVT 1525

RESULT 10
US-09-864-761-34546
; Sequence 34546, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aecmics-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
; SEQ ID NO 34546
; LENGTH: 572
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC005089.2
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 5.4
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 4
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 5.6
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.3
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.1
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 4
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.7
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3.1
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 3.7
; OTHER INFORMATION: SWISSPROT HIT: P53935, EVALUATE 2.00e-03
; OTHER INFORMATION: EST_HUMAN HIT: AW976211.1, EVALUATE 2.00e-91
US-09-864-761-34546

Query Match 36.0%; Score 2877; DB 9; Length 572;
Best Local Similarity 98.8%; Pred. No. 7.5e-154;
Matches 567; Conservative 1; Mismatches 2; Indels 4; Gaps 1;
QY 298 YMTLPSTKRNKNTGSPDRKPSKSKTNDSSLSPLNPKLWCHVHLKSLSGSPLKVKNSK 357
Db 1 YMTLPSTKRNKNTGSPDRKPSKSKTNDSSLSPLNPKLWCHVHLKSLSGSPLKVKNSK 60
QY 358 NSKSPPEHLEEMKQMSPNKLTNPHIPKGPAPKPKHSDKPLKAKGRSKGIINGOKS 417
Db 61 NSKSPPEHLEEMKQMSPNKLTNPHIPKGPAPKPKHSDKPLKAKGRSKGIINGOKS 120
QY 418 TGNKSPKKGKLTPTKTKMKTLLDMAGTKQMTAPRNSGGTPTTSKPKHKLPPAALH 477
Db 121 TGNKSPKKGKLTPTKTKMKTLLDMAGTKQMTAPRNSGGTPTTSKPKHKLPPAALH 180
QY 478 LIAYYKKNKREDKRSALSCVISKTARLLSSEDRARLPEELRSVLQKRYELLEHKRWAS 537
Db 181 LIAYYKKNKREDKRSALSCVISKTARLLSSEDRARLPEELRSVLQKRYELLEHKRWAS 240
QY 538 MSEORKEYLKKREELKKLKEKAKERREKEMLERLEKQKRYEDELTKGNLPAPRLVD 597
Db 241 MSEORKEYLKKREELKKLKEKAKERREKEMLERLEKQKRYEDELTKGNLPAPRLVD 300
QY 598 TPEGLPNTLFGDVAMVVEFLSCYSGLLLPDAQYPTAVSLMEALSADKGGFLYLNRLVI 657
Db 301 TPEGLPNTLFGDVAMVVEFLSCYSGLLLPDAQYPTAVSLMEALSADKGGFLYLNRLVI 360
QY 658 LLQTLLOTLQDEIAEDYGELGKMLSEIPLTLHSVSELVRLCLRRSDVQEESEGSDTDN 717
Db 361 ---LLQTLLOTLQDEIAEDYGELGKMLSEIPLTLHSVSELVRLCLRRSDVQEESEGSDTDN 416
QY 718 KDSAAFEDNEVQDEFLEKLETSEFFELTSEBKQLIITLALCHRIILMTYSVQDHMETROQMS 777
Db 417 KDSAAFEDNEVQDEFLEKLETSEFFELTSEBKQLIITLALCHRIILMTYSVQDHMETROQMS 476
QY 778 AELWKERLAVLKEENDKKRAEKQKREMEAKNKENGKVENGLGKTRKRIVKPFPQVDT 837


```
Db 477 AELWKEARLAVLKEENDKKRAEKQKREMEAKNKENGKVENGLGKTDKKEIVKFEQVDT 536
QY 838 EAEDMISAVKSRLLAIQAOKEREIOBEREMKVL 871
Db 537 EAEDMISAVKSRLLAIQAOKEREIOBEREMKGI 570

RESULT 11
US-09-864-761-37938
; Sequence 37938, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 37938
; LENGTH: 560
; TYPE: prt
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC005074.1
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.5
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2.8
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.9
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.93
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.99
```

```
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.8
; OTHER INFORMATION: EST HUMAN HIT: AW976211.1, EVALUE 2.00e-91
; OTHER INFORMATION: SWISSPROT HIT: P53935, EVALUE 2.00e-03
US-09-864-761-37938

Query Match 35.5%; Score 2835; DB 9; Length 560;
Best Local Similarity 99.1%; Pred. No. 1.7e-151;
Matches 559; Conservative 0; Mismatches 1; Indels 4; Gaps 1;

QY 305 TTKKNTGSPDRKPSKSKTDNSLSLSPINPKLWCHVHLKKSLSGSLKVKNSKNSKSP 364
Db 1 TTKKNTGSPDRKPSKSKTDNSLSLSPINPKLWCHVHLKKSLSGSLKVKNSKNSKSP 60
QY 365 HLEEMMKMSPNKLHTNFHIPKGPAPKPGKHSKPLKAKAGRSKGLNGKQKSTGNSKSP 424
Db 61 HLEEMMKMSPNKLHTNFHIPKGPAPKPGKHSKPLKAKAGRSKGLNGKQKSTGNSKSP 120
QY 425 KKGLKTPKTKMQMTLLDMAKGTQKMTAPRNSGGTPTSSKPKHLPALHLIAYYKE 484
Db 121 KKGLKTPKTKMQMTLLDMAKGTQKMTAPRNSGGTPTSSKPKHLPALHLIAYYKE 180
QY 485 NKDREDKRSALSCVISTARTLLSSEDRARLPBELRSIVQRYELLEHKKRWASMSSEQRK 544
Db 181 NKDREDKRSALSCVISTARTLLSSEDRARLPBELRSIVQRYELLEHKKRWASMSSEQRK 240
QY 545 EYLKKEELKKLKEKAKEREKEMLEKQRYEDQELTGKNIPLAFRLVDTPEGLPN 604
Db 241 EYLKKEELKKLKEKAKEREKEMLEKQRYEDQELTGKNIPLAFRLVDTPEGLPN 300
QY 605 TIFGDVAMVVEFLSCYSGLLLPDAQYPTITAVSLMEALSADKGGFLYLNRLVILLQ 664
Db 301 TIFGDVAMVVEFLSCYSGLLLPDAQYPTITAVSLMEALSADKGGFLYLNRLVILLQ 356
QY 665 TLLQDEIAEDYGEGLMKLSEIPLTHSVSELVRLCLRRSDVQSESGSDTDDNKDSAAFE 724
Db 357 TLLQDEIAEDYGEGLMKLSEIPLTHSVSELVRLCLRRSDVQSESGSDTDDNKDSAAFE 416
QY 725 DNEVQDEPLEKLETSEFFELTSEEKLIQILTALCHRLTMTYSVQDHMETQQMSAELWKER 784
Db 417 DNEVQDEPLEKLETSEFFELTSEEKLIQILTALCHRLTMTYSVQDHMETQQMSAELWKER 476
QY 785 LAVLKEENDKKRAEKQKREKEMBAKNGKNGKVENGLGKTDKKEIVKFEQVDTAEADMIS 844
Db 477 LAVLKEENDKKRAEKQKREKEMBAKNGKNGKVENGLGKTDKKEIVKFEQVDTAEADMIS 536
QY 845 AVKSRLLAIQAOKEREIOBEREMK 868
Db 537 AVKSRLLAIQAOKEREIOBEREMK 560

RESULT 12
US-09-839-479-1
; Sequence 1, Application US/09839479
; Publication No. US20020039779A1
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042002
; CURRENT APPLICATION NUMBER: US/09/839,479
; CURRENT FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1674
```

```
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-839-479-1

Query Match      11.7%; Score 930.5; DB 9; Length 1674;
Best Local Similarity 23.2%; Pred. No. 1.3e-43;
Matches 397; Conservative 263; Mismatches 532; Indels 521; Gaps 67;

QY 3 PLGGRKPF-----PLVNPFGPEPPTTPIHTQAFRTREYEYEARLERYSERIWTCKSTGSS 58
DB 152 PLLHRRKPFVRQPPADLPDEEVFY-CKVTNEIFRHYDDFFERTILCNLSWCAVTRGP 210
QY 59 QLTKEAWEEEOEVAELKEEPPAWYKVLVEMVH-----HNTASLEKLVDTAWLEIMTKY 114
DB 211 GLTYQEALESEKKAQNL-QSEF---EPLIIPVLVLTSLTHRSRLHEICDDIFAYVKDQY 266
QY 115 AVGECDEFEVGKMKLVKVIKHPLEKVEDEATEKSGDCADSPSSDKENSSQIAQDHQ 174
DB 267 FVEETVEVIRNNGARLQCTILEVP-----PS-----HQNGFANGHV 303
QY 175 KK---ETVVKEDEGRRESINDRARRSPKLPSTLKKGKRWAPPKFLPKYDV-----KLQ 227
DB 304 NSVDGETIISDDSETQS-----CSFQNGKKKDAIDPLL-FKYKVQPTKKEL 351
QY 228 NEDKIISNPADSLIRTPPNKEIVRVPIRHNALRAGTGENAPWVDELVKYSLPS- 286
DB 352 HESAI---VKATQISRRKHLFSRDKLFLKQHC-----EPQEGVIK---IKASSLSTY 399
QY 287 -----KFSDFLDPPKYMTLPSTTKRQNTGSPDRKPSK---SKTDNSLSGSLPNPLMC 338
DB 400 KIAEQDFSYFFDDPPTTFISANRRG-----RPPKRIHISQDN----- 440
QY 339 VHVLKXSLSGSLPKVKNKSNKSPSEHLEEMKMKSPNKLHTNPHIPKGGPAKPGKHS 398
DB 441 -VANKOTLA-----SYRSKATKER----- 458
QY 399 DKPLAKGRSKGILNGQSTGNSKSPKGLKTPTKMKQMTLLDMAKGQKMTAPRNSG 458
DB 459 DKLLK-----QSEMSLAFAKALKR----- 479
QY 459 GTPRTSSKPHKLPPAALHLIAYYKENKDREDKRSALSCVISTARLLSSEDRAKLPEEL 518
DB 480 -----EKADALEAKKEKEDKEKR-----BEL 502
QY 519 RSLVQKRYELLEHHKRWASMSBEOKEVYKKREELKKLKEKAKERREKEMLERLEKQK 578
DB 503 KKIVBEE-----RLKKKEKERLKVEREKEREKIRE--EKRYKVEYLKQMSKRP 549
QY 579 RYEDQELTG-KNLPAPFLVDTPEGLPNTLFGDVAMVBFSLSCYSGLLLPDAQYP--ITAV 635
DB 550 --EDMECDLKLPEPTPKVT--RLPPEIFGDALVLEFLNAFGLFDLODFPFGVTLE 605
QY 636 SLMEAL-SADKGGFLYLNRLVILLQTLQLL--ODEIAE----DYGELGMKLEIPIIT 688
DB 606 VLEEAALVGNDSG--PLCELLFFLTATPAQIAEVEEVAKEQLTADTKGCSLSKSLDLID 663
QY 689 LHSVSELVRLCURSDVQESGSDT-----DNKDSAAFEDNEVDQEFLEKLETSEF 741
DB 664 SCTLSEILRLHTLASGADVTSANAKRYQKRGGFDATDDACMELSLNSPLVKKLSSTSV 723
QY 742 FELTSEEKLOILTALCHRTILMTYSVQD-----HMETRQMSA 778
DB 724 YDLTGERKMKILHALCGKLLTLVSTRDFIEDYVDILROAKQEFRELKAEQHRKEEBAEA 783
QY 779 ELWKERLAVLKEENDKRAEKQKREMEARN-----KENGKVENGLGKTRDKRQIV 829
DB 784 RIRKKEEKLEKQEQKWKQKQKLEDEQORNSTADISIGEEREDFDTSIESKDETKQ-- 841
QY 830 KEPQVDTEADMIKAVKSRLLATQAK-----KERETQ-----REMKVKLRQA-E 876
DB 842 ELDDQMFTEDEDDPGSHKRGK-----RGRKGQNGFKFTRQEQINCVCNTRELLTADSEALK 897
QY 877 EBRIRKHAABEAKAFQEGIAKAKLVMRRTPIGTRNHNRYLWFSDEVPLGFIEK---GWV 933
```

```
DB 898 QEHQKEKELLEK-IQSAIACTNIF-----PLGRDMRYRYWIF-PSIPGLFIEEDYSGLT 951
QY 934 HDSI-----DYRFNHHCKDHTVSGDEDYCPSPKSKAN-LGKNASMTQHGTAATEVAVEYTT 987
DB 952 EDMLLPRSSPQNNVQSDPQVS-----TKTGEPLMSESTSNIDQG-PRDHSVOLPK 1002
QY 988 PKQGNLWFLCDSOKELDELNCLHPQGIRESQKLERL--EK-----RYQDIHHSIH 1037
DB 1003 PVHPRNRCWFYSSCEQLDQLEALNSRGRHSALKETILQESKRICAQLARFSE--EKFH 1060
QY 1038 LARKPNLGLK-----SCDGNQ-----ELLNFLRSDLIEVATRLQKGLGYVEBT- 1081
DB 1061 FSDKPPQDPKPTYSRGRSSNAVDSQMAEQLELRDLFLDIEDRIVYQGTGAIKAVTD 1120
QY 1082 -----SEFEA---RVISLEKL-----KDFECVIALQASVITKKLQFGMAPQKRR 1124
DB 1121 RHIMRSALSGRYELLSEENKENGIIKTWNEDVEIMEIDEQTKVIVK--DRLLGKITPT 1178
QY 1125 KLQSEDSAKTEVDE-----EKQVVEAKVAS-----ALEKWKTAI 1160
DB 1179 STVTNASTPGSVSVHYLAMALFOIQGIERFLKAPLDASDSGRSYKTVLDRWRBSL 1238
QY 1161 REAQTFSRMHYLLGLMDACIKWMSAENARCVCPKGEDDKLILCDECNKAFHLFCCLR 1220
DB 1239 LSSASLSQVFLHLSTLDRSVIWSKILNARCKICRKGDAENMVLCDGDRGHHTYCVRP 1298
QY 1221 ALYVDPDEWQCPACQATARNRGRNYTESASEDESDDEDEE-----EEEEEE 1276
DB 1299 KLKTVPEDGMFCPECRPKQRCRLSFRORPSLESDVEDYSDMGDEDDVDGDEEGQSEE 1358
QY 1277 EDYEV-----AGLRLRPKTI-----RGKHSVIPPAARSGRRPGKXPHSTR 1318
DB 1359 EYEVEQEDDSQEBEEVSLPKRGRPVRLPVKTRGKLSFSSFSRQGOQEPGRYPSRQQ 1418
QY 1319 SQPK-----APPVD-----DAEVDL----- 1334
DB 1419 STPKTTSSTKGRSLRKINSAPPTETKSLRIASRSTRSHGHPLOQADVVELLSPRKRRG 1478
QY 1335 -----VLQTKSSRRQSL----- 1347
DB 1479 RKSANNTPENSFPNFPNFRVIATKTSQESRSRNIASIKLSQESKRCRKRQSPSPVPT 1538
QY 1348 -----ELQCEEILHKIVKYRFSWPPREPVTDRBAEDYDVITHPMDFTQVQ 1394
DB 1539 LGRSSRGQGVHLSAFEQLVVELVRHDDSWFPLKLVSKIQVPDYDIKKPIALNIIR 1598
QY 1395 NKSCGSYRSVQEFITDMKQVFTNAEVNCRGS 1427
DB 1599 EKVNKCEYKLASEFIDDIELMFNSCFEYNPNT 1631
```

RESULT 13

```
US-10-376-537-1
; Sequence 1, Application US/10376537
; Publication No. US20030224405A1
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042001
; CURRENT APPLICATION NUMBER: US/10/376,537
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US/09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JF98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
```



```
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 69
; LENGTH: 1673
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-839-479-69

Query Match      11.5%; Score 919; DB 9; Length 1673;
Best Local Similarity 22.9%; Pred. No. 5.6e-43;
Matches 393; Conservative 262; Mismatches 534; Indels 526; Gaps 64;

QY 3 PILGRKPF---PLVNPFGCEBFFTPHTQEAFTREYEAFLERYSRIMTCKSTGSS 58
   ||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 152 PLLHRRPFVRQPPADLRDEEVFY--CKVTNEIFRHYDDFFERTILCNSLVMSCAVTGRP 210
   ||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 59 QLTKEAWEQEVAELLKEEPAWYKLVLENVH---HNTASLEKLVDTAWLEMTKY 114
   ||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 211 GLTYQALESEKARQNL--QSPF---EPLIIPVLYLTSLTHRSRLHEICDDIFAYVKORY 266
   ||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 115 AVGEEDCFVGKEKMLKVIKIHPLEKVDDEATEKKSDGACDSPSSDKENSQIAODHQ 174
   ||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 267 FVEETVEVRNNGARLOCTILEVLP-----PS-----HQNGFANGHV 303
   ||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 175 KK---ETVVKDEGRRESINDRARRSPKLTPTSLKKGKRWAPPKFLPHKYDV---KLQ 227
   ||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 304 NSVDGETIIISDDSETQS-----CSFQNGKKDAIDPLL--FKYKVQPTKKEL 351
   ||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 228 NEDKLIISNPASLSLIRTEPPNKEIVRYFIRNALBAGTGENAPWVVEDELKVKYSLPS- 286
   ||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 352 HESAI---VKATQISRRKHLFDRDKLFLKQHC-----EPQEGVIK---IKASSLSTY 399
   ||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 287 -----KFSDFLLDPYKYMTLNPSTKRKNTGSPDRKPSKK---SKTDNSSLSPNPKLWC 338
   ||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 400 KTAEQDFSVFFDDPTTFISFANRRG---RPPKRIHISOEDN----- 440
   ||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 339 VHVLKXSLGSLPKVKNXSNKSPBHELEMMKMSPNKLTNFIHPKXGPPAKKPGKHS 398
   ||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 441 -VANKQTLA-----SYRSKATKER----- 458
   ||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 399 DPLXAKGSRKGLNGCKSTGNSKSPKGLKTPKTKQKQMTLLDMAKGTQKMTAPRNSG 458
   ||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 459 DXLLK-----QEBMSKLAFAKAKLKR----- 479
   ||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 459 GTPRTSSKPKHLLPAAHLIAVYKENDKREDKRSALS CVISKTARLLSSDDRARLPEEL 518
   ||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 480 -----EKADALEAKKEKEDKPKR-----EEL 502
   ||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 519 RSLVQKRYELLEHKRWASMSBEQKEYLYKCKREELKKLKBKAKERREKEMLERLEKQK 578
   ||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 503 KKIVEE-----RLKKEEKERLKVEREKEREKLEEE-----KRKYVEYKQWSEK 546
   ||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 579 RYDEDELTG--KNLPAPRLVDTPEGLPNTLFGDVAMVVEFLSCVSGLLLPDAQVP--ITAV 635
   ||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 547 PREDMECDLDELPEPTPVKT--RLPPEIFGDALMVLEFLNAPGELFLQDFEPDGVITLE 604
   ||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 636 SLMEAL--SADKGFLYLNRVALLVLTLLQTLA--QDEIAB---DYGELGMKLSIPLT 688
   ||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 605 VLEBALVGNDSG--PLCELLPFFLTAFQATAEVEEVAKEQLTDADTKGCSLSKSLDLD 662
   ||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 689 LHSVSELVRLCLRRSDVQSESGSDTD-----DNKSOAAFPEDNEVQDFLEKLETSEF 741
   ||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 663 SCTLSEILRLHILASGADVTSANAKRYQKRGFGFATDDACMELRLSNPSLVKKLSSTSV 722
   ||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 742 FELTSEKQIITALCHRIIMTYSVQD-----HMETQQMSA 778
   ||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 723 YDLTPGCKMKILHALCGKLLTLVSTRDFIEDYVDILRQAKQBFRELKAPQHRKEREAAA 782
   ||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 779 ELWKERLAVLEENDKRAEKOKKEMEAKN-----KENGKVENGLGKTDKTRIV 829
   ||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 783 RIRKRKEELKEQEQMKKEQKLEKDEQORNSTADISIGEEREEDFDTSIESKDETEQ-- 840
   ||||| : : : : : : : : : : : : : : : : : : : : : : : : : :
```

RESULT 16

```
US-10-378-537-70
; Sequence 70, Application US/10376537
; Publication NO. US20030224405A1
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042001
; CURRENT APPLICATION NUMBER: US/10/376,537
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US/09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
```

```
QY 830 KFEPOVDTEAEDMISAVKSRRLAIQAK-----KERBIOE-----REMKVKLEQA-E 876
   : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 841 ELDQDMFTEDDDPGSKRGR-----RGKRGQNGFKEFTREQINCVTRELLTADEBEALK 896
   : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 877 EERIRKHAACAQKAFQEGITAKALVNRTPIGTDNRHNNYWLFSDEVPGLFIEK---GWW 933
   : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 897 QEHQRKEKELLEK--QSAIACTNIF-----PLGRDRMYRRYWF--PSLPLGLFIEEDVSGLT 950
   : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 934 HDSI-----DYRFNHCKDHTVSGDEDCYCPRSKAN--LGNKASMTQHGTTATVAVVETTT 987
   : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 951 EDHLLPRSPSFQNNVQSQDPQVS-----TKTGEPLMSESTSNIDQG--PRHDSVQLPK 1001
   : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 988 PKQGNLWFLCDQSQKELDELLNCLHPQGIRESOLKERL----- 1025
   : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1002 PVHKPNRMCFYSSCEQDLIEALNSRGHRESALKETLLQEKSRICAQLARFSEEFHFS 1061
   : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1026 EKRYODIHSIHILARKPNLGLSKCQNGQ-----ELNLFRLSRLIEVATRLQKGLGVVEE 1080
   : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1062 DKRQDPDSKPTYRGRSSN-----AYDPSQMCABKQLELRDLFLDIEDRIYQGTGLGAIKV 1117
   : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1081 T-----SEFEA---RVISLEKL-----KDFGECVIALQASVIKKFLOGFMAPKQK 1122
   : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1118 TDBHWRSALESGRYELLSEENKENGIIKTVNEDEVEMEIDEQTKVIVK--DRLLGIKTE 1175
   : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1123 RRLQSEDSAKTEEVDE-----EKQWBEAKVAS-----ALEKWK 1158
   : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1176 TPSTVSTNASTPQSVSSVVHYLAMALFQIEQGLERRFLKAPLDASDSGRSYKTVLDRWRE 1235
   : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1159 ATRQAQTSRMVHLGMLDACIKWDSANARCKVCPKGEDDKLLILCECNKAPHLFCL 1218
   : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1236 SLSSASLSQVFLHLSTLDRSVIWSKILNARCKICRKGDAMNMYLDCDGRGHHTYCV 1295
   : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1219 RPALEYVPDGEWQCPACOPATARRNSRGRNYTEESASEDESEDESEEE---EEDDEE 1274
   : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1296 RPKLKIIVPEGDFCEPCRPQCRRLSFRQPSLESDDEVDSMGWGEDDEVDEDEEGQS 1355
   : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1275 EEDYEV-----AGLRLRPRTI-----RGKHSVIPPAAARSGRRPGKXPHST 1316
   : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1356 EEEVEVEQEDDSQEEBEVSLPKRGPQVRLPVKTRGLKSSFSRSGQOQEPGRYPSRS 1415
   : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1317 RRSQPK-----APPVD-----DAEVEL----- 1334
   : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1416 QOSTPKTTSVSKTSLRKINSAPPTETKSLRIASRSTRHSHGHPLOADVVELLSPRKR 1475
   : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1335 -----VLQTKRSSRSQSL----- 1347
   : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1476 RGRKSANNTPENSPNFPNFRVIATKSSQSRSVNIATSKLSQESSESKRCKRQSPSP 1535
   : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1348 -----ELQKCEBILHKIVKYRFSWFPREPVTREADEADYDVIHPMDFQT 1392
   : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1536 VTGLRRSSRGQGVHLSAFEQLVVELVRHDDSWPFLKLVSKIQVDPDYDIKKPIALNI 1595
   : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1393 VONKSCSGSYRVSQVFLTDMKQVFTNAEVYNCGRS 1427
   : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1596 IREKNVCKEYKLASEFIDDIELMFNSCFEYNPRNT 1630
   : : : : : : : : : : : : : : : : : : : : : : : : : :
```

```
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 70
; LENGTH: 1673
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-376-537-70

Query Match      11.5%; Score 919; DB 15; Length 1673;
Best Local Similarity 22.9%; Pred. No. 5.6e-43;
Matches 393; Conservative 262; Mismatches 534; Indels 526; Gaps 64;

Qy 3 PLLGRKPF-----PLVNPLPGSEPPPTIHTQEAFTREYEARELRYSERIWTCKSTGSS 58
Db 152 PLLHRKPVROKPPADLRDEVEFY-CKVTNEIFRHYDDFFERTILCNLSVMSCAVTGRP 210
Qy 59 QLTKEAWEEQEVAEALLKEFPWYKLVLEMVH---HNTASLEKLVDTAWLEIMTKY 114
Db 211 GLTQEALESEKARQNL-QSFP---EPLIIPVLVLTSLTHRSRUHEICDDIFAYVKORY 266
Qy 115 AVGECDFEVEGKMLKVIKVIHPLEKVEBEATEKSDGACDSPSSDKENSQIAQDHQ 174
Db 267 FVEETVEVRNNGARLQCTILEVLP-----PS-----HQNGFANGHV 303
Qy 175 KK---ETVVKEDEGRRESINDRARSPKLPSTLKKGERKWAPPKFLPHKVDV-----KQ 227
Db 304 NSVDGETIISDSDSETQS-----CSFQNGKKDAIDPLL-FKYKVQOPTKKEL 351
Qy 228 NEDKIISNVADSLRTERPPNKEIVRYFIRHNALRAGTGNAPMVVDELVKKYSLPS- 286
Db 352 HESAI---VKATQISRRRHLSFRDKLFLKQHC-----EPQEGVIK---IKASSLSTY 399
Qy 287 -----KFSDFLLDPYKMTLPSTPKRKNYTGSPDRKPKK---SKTDNSLSPLNPKLWC 338
Db 400 KIAEQDFSYFFDDPTTIFSPANRRG-----RPPKRIHISQEDN----- 440
Qy 339 HVHLKKSLSGSLPKVYNSKNKSPDEHLEEMKQMSPNKLHTNHIPKKGPPAKPGKHS 398
Db 441 -VANKQTLA-----SYRSKATKER----- 458
Qy 399 DKPLKAGKSGILNGQKSTGNSKSPKGLKTPKTKMOMTLMDMAKGTQKTRAPRNSG 458
Db 459 DKLLK-----QEEMSLAFKAKLR----- 479
Qy 459 GTPRTSKPHKLPAPALHLIAYYKENKDRKRSALSCVISTARLLSSEDRALPEEL 518
Db 480 -----EKADALEAKKKEKEDKEKR-----EEL 502
Qy 519 RSLVQKRYELLEHKRWASMSSEEQKQKYLKKREELKKLKEKAKERREKEMLERLEKQK 578
Db 503 KKIIVEE-----RLKKKEERLKVEREKEREKUREE-----KRKYVEYKQWSK 546
Qy 579 RYEDQELTG-KNLPAPRLVDPTEGLPNTLFGDVAMVVEFLSCYGLLLPDAQYP--ITAV 635
Db 547 PREDMECDLKLPEPTPVKT--RUPPEIFGDMALVLEFLNAFGLFDLQDEFDPDQVLE 604
Qy 636 SLMEAL-SADKGGFLYLNRLVILLQTLQTL--QDEIAE-----DYGELGMKLSIEPLT 688
Db 605 VLEEAALVNDSEG--PLCELLFFFLTAIFAABEEVEVAKELTDAATKGCSLKSLLD 662
Qy 689 LHSVSELVRLCLRRSDVQEESEGSTD-----DNKDSAAPEDNEVQDEFLEKLETSEF 741
Db 663 SCTLSEILRLHLIASGADVTSANAKRYQKRGGFDTDDACMELNSPLSVKLSLSTSV 722
Qy 742 FELTSEKLOJLTALCHRLIMTYSVD-----HMETRQQMSA 778
Db 723 YDLPTEGKMKHLIHALCGKLLILVSTRDFIEDYVDILROAKQEFRELKAEQHKREBEAA 782
Qy 779 ELWKERLAVLKEENDKRAEKQKREMEAKN-----KENGKVENGLGKTKRKRIV 829
Db 783 RIRKKEEKLEQEQKWKKEKQKEDDEBQNSADISIGEEREDFDTSIESKDETK-- 840
```

```
Qy 830 KFEQVDTEAEDMISAVKSRRELLAIQAK-----KEREIQB-----REMKVKLEROA-E 876
Db 841 ELDDQMTFTEDEDDPGSHKGR-----RGKRGQNGKFEFTRQIQINCVTRELTADAEEALK 896
Qy 877 BERIRKHAATAEAFQEGIAKAKLVMRRTPIGTDRNNHRYLWLFSDDEVFGLFIEK---GMV 933
Db 897 QEHQKEKELEK-LQSAIACTNIP-----PLGRDRMYRRYWF-PSLPGLFIEDYSGLT 950
Qy 934 HDSI-----DYRFNHCKDHTVSGDEDECYPRSKKAN-LGKNASMTQGTATATEVAVETTT 987
Db 951 EDHLLPRSPSPQNNVQSDPQVS-----TKTGEPLMSESTSNIDOG-PRDHSVOLPK 1001
Qy 988 PKQONLWFLCDSOKELDELINCLHPQIGIRESQKLERL----- 1025
Db 1002 PVHPKRWCFYSSCEQDOLLEALNSRGHRESALKETLLQEKSRICAQLARPSEBKFPFS 1061
Qy 1026 EKRYQDIIHSHLARKPNLGLKSCDGNQ-----ELNLFNRSDLIEVATRLQKGGLYVEE 1080
Db 1062 DKRQDPSKPTYSRGRSSN-----AYDPSQMAEKQLLELRDLFLDIEDRIYQGTLAGIKV 1117
Qy 1081 T-----SEFEA---RVISLEKL-----KDFGECVIALQASVKKFTLOGFWAPKQ 1122
Db 1118 TDRHWRSALESRGYELLSEENKENGIIKTVNEDEVEMEIDEQTKVIVK--DRLLAGIKTE 1175
Qy 1123 RRKLQSEDSAKTEEVDE-----EKMVVEEAKVAS-----ALEKWK 1158
Db 1176 TPSTVTNASTPQSVSSVVHYLAMALFOIEOGLERRFLKAPLDASDSGRSYKTVLDWRRE 1235
Qy 1159 AIRBAQTPSRMHVLLGLMDACIKWDMSAENARCKVCPKKGEDDKLILCDECNKAPHLFCL 1218
Db 1236 SLLSSASLSQVFLHLSTLDRSVIWSKSLNARCKICRKKGDAENMVLCDGCDRGHTTYCV 1295
Qy 1219 RPALYVEPDGWEQCPACOPATARNRSGRNTYESASESDSDSDSDDEEBE-----BEEEE 1274
Db 1296 RPKLIVPEGDWPCPCRPKQRCRLSPRQPSLESDESDSDSDMGEDDEVDGDEEGQS 1355
Qy 1275 EEEYDEV-----AGLRLRPKTI-----RGKHSVIPPAARSRGPKKPHST 1316
Db 1356 EEEYEVEQEDDQSEEEVSLPKRGRQVRLPVKTRGLSSSSFSRQQQEPGRYPERS 1415
Qy 1317 RRSQPK-----APPVD-----DAEYDEL----- 1334
Db 1416 QOSTPKTTVSSKTSRLRKINSAPPTKSLRIASRSTRHSHGHPQLQADVVELLSPRKR 1475
Qy 1335 -----VLQTKRSSRRSL----- 1347
Db 1476 RGRKSANNTPENSPNPFNFRVIATKSSQSRSVNIASKLSLQESKSRCKRQSPSPSP 1535
Qy 1348 -----ELQKCEEILHKIVKRYFSPREPVTDRDEADYDVVITHPMDFTQ 1392
Db 1536 VTLGRSSGROGGVHELSAFEQLVVELVRHDDSPFLKLVSKIQVDPDYDIKKFIALNI 1595
Qy 1393 VQNKCSGYSRVSQVFFLTDMKQVFTNAEVYNCRGS 1427
Db 1596 IREKNVKEYKLASEFIDDIELMFNCFEYXPRNT 1630
```

RESULT 17

```
US-10-702-148-69
; Sequence 69, Application US/10702148
; Publication No. US20040063145A1
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042002
; CURRENT APPLICATION NUMBER: US/10/702,148
; CURRENT FILING DATE: 2003-11-05
; PRIOR APPLICATION NUMBER: PRIOR APPLICATION NUMBER: US/09/839,479
; PRIOR FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
```



```
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1134
; LENGTH: 141
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (105)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-864-1134

Query Match      8.9%; Score 712; DB 9; Length 141;
Best Local Similarity 99.3%; Pred. No. 1.2e-32;
Matches 140; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 464 SSKPHKHPAALHLIAYKENKREDKRSALSCVISKTARLLSSDRARLPPEELRSVLQ 523
Db 1 SSKPHKHPAALHLIAYKENKREDKRSALSCVISKTARLLSSDRARLPPEELRSVLQ 60

QY 524 KYELLEHKRWASMSSEQRKEYLKKREELKKLKEKAKERREKEMLERLEKQRYEDQ 583
Db 61 KYELLEHKRWASMSSEQRKEYLKKREELKKLKEKAKERREKEMLERLEKQRYEDQ 120

QY 584 ELTGKQLPAFLVDTPEGLPN 604
Db 121 ELTGKQLPAFLVDTPEGLPN 141

RESULT 19
US-10-087-192-663
; Sequence 663, Application US/10087192
; Publication No. US20020182586A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
; FILE REFERENCE: 529452000122
; CURRENT APPLICATION NUMBER: US/10/087,192
; CURRENT FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 2059
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 663
; LENGTH: 1586
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-087-192-663

Query Match      7.5%; Score 599; DB 13; Length 1586;
Best Local Similarity 20.9%; Pred. No. 5.6e-25;
Matches 318; Conservative 231; Mismatches 549; Indels 422; Gaps 61;

QY 137 IHPLSKVDEATE--KXSDGACDSFSDXE-----NSSQIAQDHQKKTETVYKE 182
Db 236 IHPEAAEKELTSVVAENGTGLVLSLELEEEQPELKMCGYNGSVSVESLHQSVLVDP 295

QY 183 -----DEGRRESINDRARRSPKLPSTLSKKGKRWAPKFLPHKY--DVKLQN-EDKI 232
Db 296 PTVSCIDD--PSHLPDQLEDLILSEDSLEPPDSLAAAEFVSGSLYGIDDAELMGAEDKL 353

QY 233 -----ISNVPADSLIRTPPNKEIVRYFIRHNALRAGTGENAPWVVERDE 277
Db 354 PLEGNPVISALDCPALSNANAFSLADDQSQTASIF-----VSPTSPPVLGES 401

QY 278 LVYKYSLSKFSDFLLDPKYNTLNPSTYKRWGTGSPDRKPSKSKTDSNSSLSPNPKLW 337
Db 402 VLQGSPEFSPAAAF-----QTVSPA--RKNVSSAPKARADREETGGAVA----- 444

QY 338 CHVLLKKSLSGSPKVKNSKNSPEEHLEEMMKWMSPNKLTHTNFIKP----- 386
```

```
Db 445 -----VSGSDVLK-----RRIATPEZ-----VRLPLQHGWRREVRIKKSGSHWQGETWY 489
QY 387 KGPPAKPGKHSKPLKAKGRSKGILNGOKSTGNSKSPKGL-----KTPK-TYKMQM 438
Db 490 YGPGCKRMKF---PEVIKYLRSNVVHSVRREHFSFSPMPVGDFFEDRTPEGLQWQL 546
QY 439 TLLDWAQTKQWTRAPRNSGGTPTSSK-PHKHLP-----PAALHLIAYKENKRE 489
Db 547 SAEIIPSRIOAIT---GKRGRPRNNEKAKNKEPVKVRGRGRPPPKIKPELLNKTDRLL 602
QY 490 DKRSALSCVISKTARLLSSDRARLPPEELRSVLQRYELLEHKRWASMSSEQRKEYLKK 549
Db 603 PKK-----LETQELSEDDKAKNTKWKQKQK-VORGESQTPVQCGARKRKQDTYS 654
QY 550 -KREELKKLKEKAKERREKEMLERLEKQRYEDQELTG-KNLPAFLVDTPEG--LPN 604
Db 655 LKQDTKKKLKEEQORQAILE--EMKXPTEDMCLSDHQPLPDFTRI---PGLTLLS 709
QY 605 TLFGDVAMVVEFLSCYSGLLLPDAQYPITAVS-LMEAL-----SADKGGFLYLNRLVIL 658
Db 710 RAFSDCLTIVEFLHSFGKVLGFDLTQVPSLGLVQEGLLCOGDSLDK-----VQDL 760
QY 659 LQTLLOTLQDEIAEDYGE---LQMKLSEIPLTLHSVSELVRLCLRRSDVQEESEGST 714
Db 761 LVRLKKAALHDPGLPPYQCSLKILGEKMSIPLTRDNVSEILRCFLMA----- 808
QY 715 DDNKDSAAFEDNEVQDEFLEKLETSFFELTSEBKLIQILTALCHRI----LMTYSVQDHM 770
Db 809 -----YRVEPSCDLSLRTQPPFOAQQPQKAAILAFVLVHLSSTIINEIDKTL 857
QY 771 ETRQOMSALW--KERLAVLKEENDKKAERKQKEMEKAKKENGKENGKENGKENGKENGK 828
Db 858 ESVSSCRKNKWIVEGLRLRLTALAKRTGRPEVWME-----GAEDGLGRR-RSSRI 907
QY 829 VKFEPQV--DTEAEDMISAVKGRRLLAIOAKKEREIOEREMKV-KLEROAEEERIRKHA 886
Db 908 MEETSGIEEEEEENTTAVHGR-----GRGEIDVAASSIPELERHIEKLSK----- 957
QY 887 AKAPQEGIAKAKLVMRTPIGTDNRHNRYWLFSDVEPGLFIE--KGWVHSDIDYRNNH 944
Db 958 -QLFPRKLLHSSQMLRAVSLGQDRYRRHYWL-PYLAGIFVGESEGSTGPFWNFSAQIP 1015
QY 945 C-----KDHTVSGDEDYCPRSKANILGNASMT----- 973
Db 1016 CDAAPTPTPPPAVSEDO---PTPSLQALLASSKPMNTPGAANPCSPVOLSTHLPDGGTPKRLS 1072
QY 974 -----OHGTATEVAVVETTPKQGNLWFLCDSQK 1002
Db 1073 GDSEMSQSPGTLGQPKRGRPPSKFFKQVQEH-YLTQLTQAPIPEPCSGWVWIRDPE- 1130
QY 1003 ELDELLNCLHPQGIRESOLKERLEKRYQDIITHSHLAKPNLGLKSCDGNQELLN--FLR 1060
Db 1131 TLDVLLKALHPRGIREKALHKHLSK-HKDFLQEVCL-----QPLTDPIPEP 1175
QY 1061 SDLTEVATRLQKGGIYVEETSEFEARVISLEKLDKFGCEVIALQASVIKKFLOGFMAP- 1119
Db 1176 NEL----PALBEGVMSWSPKETYETDLAVLQWVEELEQRVWLSLQI-----RGWTGPT 1226
QY 1120 -----KQKRRKIQ 1127
Db 1227 PDSTREDITYCEHLPDSPEDIPWRGRREGTVPQRQNNPLDLAVMLAVLEQNVERRYL 1286
QY 1128 SEDSAKTEVDEEEKOWE-----EAKVASALEKWKTAIRSAQTFSRMHVLIG 1174
Db 1287 REPLMAAEVVVVEKALLSTPNAGPDGTSTSEIYETPRVVRWROTLCRCSAAQCLCMG 1346
QY 1175 MLDACIKWDMANARCKVCKGDDDKLILCDECNKAFHLFCLRPALYEPDVGWQCPA 1234
Db 1347 QLESIAMEKSVNKVTCILVCRKGNDEFLLCDGCRGCHYCHRPKMEAVPEGWFCAV 1406
QY 1235 QPATARNRNGRNYTEESASEDSDDSDSEEEEEEEEDYEAGLRRLRPRTKING 1294
```

Db 1407 CL-----SQVEBYTORPGPKGQK----- 1427

Qy 1295 KHSVIP---PAARSGRRPGKPHSTRASQKAP--PVDDAEVDVLQTKRSSRRQSLEL 1349

Db 1428 RKSFPPLTPPEGDSRRR---MLSRSDSPAVPRYPEDGLSPK---RRHSHMRSHSDL 1480

Qy 1350 QKCEIILHKVYRFSPPREPTRDEAEYDYVITHPMDPQTVQNKCSGYSRVSQEF 1409

Db 1481 TFCIILMEMESHDAWPFLFEPYVNLVSGYRVRVKNPMDFSTMRERLLGGYTSSEFA 1540

Qy 1410 TDMKQVFTNAEVNCRGSHV 1429

Db 1541 ADALLVFDNCQTFNEDDSEV 1560

RESULT 20

US-09-839-479-21

; Sequence 21, Application US/09839479

; Publication No. US20020039779A1

; GENERAL INFORMATION:

; APPLICANT: Jones, Michael H.

; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR

; FILE REFERENCE: 06501-042002

; CURRENT APPLICATION NUMBER: US/09/839,479

; CURRENT FILING DATE: 2001-04-20

; PRIOR APPLICATION NUMBER: US 09/418,710

; PRIOR FILING DATE: 1999-10-15

; PRIOR APPLICATION NUMBER: PCT/JP98/01783

; PRIOR FILING DATE: 1998-04-17

; PRIOR APPLICATION NUMBER: JP 9/310027

; PRIOR FILING DATE: 1997-10-24

; PRIOR APPLICATION NUMBER: JP 9/116570

; PRIOR FILING DATE: 1997-04-18

; NUMBER OF SEQ ID NOS: 72

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 21

; LENGTH: 1972

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-839-479-21

Query Match 6.7%; Score 533; DB 9; Length 1972;

Best Local Similarity 18.0%; Pred. No. 3.8e-21;

Matches 353; Conservative 256; Mismatches 598; Indels 758; Gaps 68;

Qy 8 KPPLVNLPGEBPF-FTIPH-----TOEAFRTREEVEARLERYSERIWTCKST 55

Db 189 KPULVNOAKETYMKLIVPSPDLKAGNKTSESSLITSELRKREQ-----KOA 241

Qy 56 GSQLTHKEAWEEQEVALLK-----EBFPWYKLVLEMVHHN-----TA 97

Db 242 FPSQLKKQESSKSLKVIAALSNPATSSSPA-HPKQTLNHNHPNPLTNALLGNHPNG 300

Qy 98 SLEKLVDTAWLEMTKYAVEECDFEV-----124

Db 301 VIQSIOEAPLALTTTKMQSKINENIAAASSTFFSPVNLSTSGRRTPGNOTPVMPSPAS 360

Qy 125 -----GKEKML--KVIVKI-----HPL-----EKVDEE 146

Db 361 PILHSGCKEAVNNVNPVKTQHSHPAKSLVQFPGTSDIPSSKDSNDEDEE 420

Qy 147 ATEKSDGACDSPSSDKENSQIAODHQKQKTVVKEDEGRRESINDRARRSPKRLPTSLK 206

Db 421 EDEEDEDDESDDSQESDSNSESDETEGSEEDDDDKQDESDD-----T 467

Qy 207 KGERKWAPPKFLPHKYDVVKLQNEDKIISNV--PADSLITERPPNKEIVYFIRHNALRA 264

Db 468 EGE-----KTSMKL---NKTTSVSKSPMSLTGHSTPRNLHIAK-----503

Qy 265 GTGENAPWVEDELVKYSIPSKFSDFLDPKYKMTLNPSTKXKNKTSPPDRKPSKSK-T 323

Db 504 -----APGSAPALCESQSPA-----FLGTSSSTL---TSSPHSGTSKRRRV 544

Qy 324 DNSLSSSPLNPKLWCHVHLKKSLSGSPLVKXNSKNSKSP-----EEHLEEMKMMKSPN--- 376

Db 545 DERELRIPLFEGY-WQRETRIRNFGG---RLQGEVAYYAPCGKKLRQYPEVIKYLRSNGIM 600

Qy 377 -----KLHT--NFHI PKKGP-----PAKKPC 395

Db 601 DTSRDNFSAKIRVGDIFYEARDGPOEMQWCLLKEEDVIPRAMEGRGRPPNPORQRA 660

Qy 396 KHSIDKPLKAGRSKGT-----LNQCKSTGNS 421

Db 661 REESRMRRKGRPPNVGNNAEFLDNADAKLLRLQAEIARQAQAIKLLRLKQKQEQARVA 720

Qy 422 KSPKK--GLKTPKTKMQMTLLDMAKTOMTRAPRNSGGTPTRTSSKPHHLPPAALHLI 479

Db 721 KEAKKQQAIAAAEKRKQEQIKIMKQEQEKIRIQO-----IRMEKEL--RAQQIL 769

Qy 480 AYYKENKOREDKRSALS CVISKTARLLSSEDEARLPEELRSLVQKEYELLEHKKRWASMS 539

Db 770 EAKKKKEE-----AANAKLLEAEKKIK-EKEMR---RQOAVLLKHQER----- 809

Qy 540 EBQRKEYLKKREELKKLEKAKERREKEMLERLEKQKRYEDQEL---TGKNL----- 590

Db 810 ERRROHMLMKAEARKKAEERLQKQKDEKLNKERKLQRRRLEMAKELKXPNED 869

Qy 591 -----PAFRLVDTPPE-GLPNTLFGDVAMVVVEFLSCYSGLLLPDAQYPTITAVSLMEALS 642

Db 870 MCLADQKPLPELPRIPLGLVLSGSTFSDCLMVVQFLRNFVKGLGFDVNDVPNLSVLQE-- 927

Qy 643 ADKGGFL-----YLNRVLVILLQTLQTLQDRIAE DY-----GELGMKLSIPLTLHVSSEL 695

Db 928 ----GLLNIGDSMGVEVDLLVRLLSAAVCDPLITGYKAKTALGEHLNVGNVRNDVNSEI 983

Qy 696 VRLCLRRSDVQBESEGSDDDNKDSAAAFEDNEVQDEFLEKLETSEFFELTSEEKLOILTA 755

Db 984 LQIFM-----BAHQCTQELTESLTKTKAFQAHTPAQKASVLAF 1020

Qy 756 LCHRILMTYSVQDHMETROOMSAELWKERLAVLKEENDKQKAEKQKQKEMAKNKGKGV 815

Db 1021 LINELACSKSVVSEIDKNIDYMSNLRDRKWV---EGKLRKLRIRIHAKKTGRDTSGGI 1076

Qy 816 ENG-----LG-----KTDRKKRIVKFPEQVDTEAEDMISAVKSRRLLAIOAKKEREIOE 864

Db 1077 DLGEFOHPLGTTPGKRKRKRGDSYDDDDDDDDDD-----QGEDDEDEE 1123

Qy 865 --REMKVKLERQAEERIRKHKAAAEK-----AFQEGIAKAKLVNRRTPIGT 909

Db 1124 DKEDQKGGKTDICEDEDEGDOAASVEELEKQIEKLSKQOSYRRKLFDSHSLRSVMFGP 1183

Qy 910 DRNHNRYWLFSDVPGLFIE-----KGWVH-----DSIDY 939

Db 1184 DRYRRRYWIL-PRCGGIFVEGMESGEGLEBIEIAKEREKLLKKAESVQIKEBMFETSGDSLNC 1242

Qy 940 RFNHHC-----KDHT---VSGDEDYCPRSK-----KANLGRNA--- 969

Db 1243 SNTDCEQKEDLKEKDNNTNLFQKPGSFKLSKLEAVAKMPPESEVMTPKPNAGANGCTL 1302

Qy 970 -----SMNTQHGTEATEVAVE-----TTTPKQGO 992

Db 1303 SYONSQKLSLGSVQSTATQSNVEKADSNLFTGSSGPKGYFSPLPNDOLLKLTLEKNRQ 1362

Qy 993 NLWFL-----CD-----999

Db 1363 --WFSLLPRTCCDTSLTHADMSTASLVTPQSQPPSKSPSPTPAPLGSSAQNPVGLNPPA 1420

Qy 1000 -----999

Db 1421 LSPQVQKGVSMGLQFCGWPTGVVTSNIPFTLSVPSLGSGLGSEGNNSFLTSNVASS 1480

Qy 1000 -----SOKELDELLNCLHP 1013

Db 1481 KSESPVQNEKATSAQPAAEVAKVPDFFSPKPIPEEMQFGWIRIDPDLLKALLKVLHL 1540

Qy 1014 QGIRESQLKERLEKRYQDIIHSIHLARKPNLGLKSCDGNQELLNFRSLDIE---VATRL 1070

```
Db 1541 RGIREKALQKQK-HLDYITQACLKWK-DVAIIELNENE--NQVTRDIVENWMSVEQA 1596
QY 1071 QKGGYGVETSEFARVITSLBKLDGFCVIALQASVIKKFQGFMAPKQ-----1121
Db 1597 MEMDLSVLQQVEDLERRVAS-----ASLQ-----VKGWCPEPASEREDLVY 1638
QY 1122 -----KRRKLQSEDSAKTEVEDEEKKWVE--EAKVAS 1151
Db 1639 FEHKSFTKLCKEHGEFTGEDESSAHALERKSDNPLDIAVTLADLERNIERRIEDIAP 1698
QY 1152 ALEKWKTAIREAQTSRMHVILGMLDACIKWMSAENARCKVCPKKGEDDKLILCEGCK 1211
Db 1699 GLRWLRALSEARSAAQVACIQQLQKSIWAKSIWKVYCOICRKGDNELLLDCGCDK 1758
QY 1212 AFLHCLRLPALYVEPDGEVQCPCAPQATA-----RRNSRGRNYESASB-----DS 1258
Db 1759 GCHTYCHRPKIITIPDGDWFCFACIAKASGQTLKIKLHVKGKKTNESKKGKVTLTGDT 1818
QY 1259 EDESDSEEEEEEEEEDEYVAGLRULRPRTIRGKHSVIPPAARSGRRPGKKPHSTR 1318
Db 1819 EDEDS-----ASTSSSLKRGKDKLQKRR 1841
QY 1319 SOPKAPPVDDAEVDLVLQTKSSRRQSLQKCEIILHKIVKYRFSWPPFREPVTTRDEAE 1378
Db 1842 MEENTS--INLSKQSFSTVKKPKRDDSKDLALCSMLITEMETHEDAWPFLPVNLKLP 1899
QY 1379 DYVDVITHPMDFOTVQNKCSGYSRVQBFLLTDMKOVFTNAEYVN 1423
Db 1900 GYKVKIKPMDFTIREKSSGOYPNLETFALDVLVFDNCETFN 1944

RESULT 21
US-10-376-537-21
; Sequence 21, Application US/10376537
; Publication No. US20030224405A1
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042001
; CURRENT APPLICATION NUMBER: US/10/376,537
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US/09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 21
; LENGTH: 1972
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-376-537-21

Query Match 6.7%; Score 533; DB 15; Length 1972;
Best Local Similarity 18.0%; Pred. No. 3.8e-21;
Matches 353; Conservative 256; Mismatches 598; Indels 758; Gaps 68;

QY 8 KPFPVLNPLGPEPP-FTIPH-----TQAFRTREYEARELRYSERIWTCKST 55
Db 189 KPLSLVNQAKETYMKLIVPSDVLKAGNKNTSESSLLTSELRSKREY-----KQA 241
QY 56 GSSQLTHKAEWEQVQVAILK-----EFPRAWYKLVLEMVHNN-----TA 97
Db 242 FPSQLTKGSSKSLKKVIAALNPNKATSSSPA-HPKQTLNHNHPFLTALLGNHPNG 300
QY 98 SLEKLVDTAWLEIMTKYAYGECDFEV-----124
Db 301 VIQSVIQEAPLALTITTKKQSKINENIAAASPTPSSPVNLSTSGRRTPGNQTPWMPAS 360
```

```
QY 125 -----GKEKML--KVKIYKI---HPL-----EKVDEE 146
Db 361 PILHSQGGKEKAVNNVNPVKTHSHHPAKSLVQPGRTGSDIPSSKDSNEDEE 420
QY 147 ATEKSDGACDPSDSDKENSQIAQHQKQKQKTVVKEDEGRRESINDRARRSPKLP 206
Db 421 EDEDEDDDDSDSDSSESDTSGSEEDDDDKQDESDS-----T 467
QY 207 KGERKMAPPKFLPHKYDVKLQNEDKIISNV--PADSLIRTERPPNKEIVRVFIR 264
Db 468 EGE-----KTSMKL---NKTTSSVSPSLSLTGHSTPRNLHIAK-----503
QY 265 GTGENAPWVDELVKYSLSKPSDFLLDPKYMTLNPSTKRNKNTGSPDRKPSKSK 323
Db 504 -----APGSAALACSESQSPA-----FGTSSSTL---TSPHSGTSKRRVT 544
QY 324 DNSSLSPLNPKLWCHVHLKKSLSGSLPKVKNKSNKSP---EHLBEMKQMSPN---376
Db 545 DERELRIPLYG-WQRETRIRNFGG---RLOGEVAYYAPCGKQLQYEVIKYLSRNG 600
QY 377 -----KLHT-NFHIPKKGP-----PAKPG 395
Db 601 DISRDNFSAKIRVGDYEARDGPQEMQWCLLKEDVIPRAMEGRGRPPNDRORA 660
QY 396 KHSKPLKAKGSKGI-----LNGKSTGNS 421
Db 661 REESMRRRKGRPPNVGNAEFLDNADAKLLAKLQAEIARQAQIKLRLKQKQOARVA 720
QY 422 KSPKK--GLKTPKTKQKQMTLLDMAKGTQKQTRAPRNSGGTPTRTSSKPHKHLPP 479
Db 721 KEAKQQAIAAEKQKQEQIKIMQKQEKIKR10Q-----IRMEKEL--RAQQL 769
QY 480 AYYKENKDREDKRSALS CVISKTARLLSSEDRARLPBELRSIVQKRYELLEHKK 539
Db 770 EAKKKKEE-----AANAKLEAEKRIK-EKEMR---ROQAVLLKHOER-----809
QY 540 EQRKEYLKKREELKKLKEKAKERRREKEMLERLEKQKRYEDQEL---TGQNL-----590
Db 810 ERROHMLMKAMEARKAEKELKQEKREKLINKERKLEQORLELEMAKELKKPNED 869
QY 591 -----PAFRLVDTPE-GLPNTLFGDVAMVVFSLSCYSGLLLPDAQYPIATAVSL 642
Db 870 MCLADQKPLPELPRIPLGLVLSGSTFSDCLMVVQFLRNFQKVLGFDVNDVPNLSV 927
QY 643 ADKGGFL---YLNKRVLLQTLQTLQDEIABDY---GELGMLKSEIPLTLHSVEL 695
Db 928 ---GLLNGIDSMGEVDLLVRLLSAAVCDPLITGYKAKTALGSHLLNVGYNRDNV 983
QY 696 VRLCLRRSDVQSESESGSDTDDNKDSAAAFEDNEVDQEFLEKLETSFEFFELTSE 755
Db 984 LQIFM-----EACGQTELTESLTKAFQAHTPAKASVLAF 1020
QY 756 LCHRILMTYSVDQHMETRQMSAELWKERLAVLKEENDKQGAEKQKREMEAKNKENG 815
Db 1021 LINELACSKSVSEIDKNIDYMSNLRDRKVVV---EGKLRKLRITIIHAKTKGRD 1076
QY 816 ENG-----LG-----KTRDKKRIIVKFPQVDTEADMSIAVKSRRLLAIQAK 864
Db 1077 DLGEHQHPLGTPTPKRRRRKGGSDYDDDDDDSD-----QDEDEDEDE 1123
QY 865 ---REMKVKLERQAEERIRKHKAAAEK-----AFQEGIAKAKLVMRPTTGT 909
Db 1124 DKEDQKGTTCICEDEGDAQAAVEELEKQIEKLSKQSQYRRRLKFASHLSRVMFP 1183
QY 910 DRNHRYWLFSDVEVPGLFIE-----KGWVH-----DSIDY 939
Db 1184 DRYRRYIL-PRCGGIVFEGMESGEGLEEAETAKERELKKAESVOIKEEMFETSGD 1242
QY 940 RFNHHC-----KDHT---VSGDEDYCPRSK-----KAMLGKNA---969
Db 1243 SNTDHCQEKDLKENDTNLFLQKPGSPFKLSKLLEAVAKMPESEVMTPKPNAGANG 1302
```

Qy	970	-----SMTQHGTA TEAVE-----	TTTPKQGQ	932
Db	1303	: : : : SYQNSGKSLGSVOSTATQSNVERADSNLNTGSSGPKGYFPLPNDQLLKLTTEKNRQ		1362
Qy	993	NLWFL-----CD-----	-----	999
Db	1363	--WFSLPRPTFCDDTSLTHADMSTASLVTPQSOPPSKSPPTAPLGSSAQNPGVLNPPA	1420	
Qy	1000	-----	-----	999
Db	1421	LSPLOVKGVSMGLQFCGMPTGYVTTSNIPTLSVPSLGSGLGLESGNGNSFLTNNVASS	1480	
Qy	1000	-----	-----SQKELDLNLCLHP	1013
Db	1481	KSESPVQNEKATSQAQPAAEVAKPVDFPSKPPIPEEMQFGWRIIDPEOLKALLKVHLH	1540	
Qy	1014	QGIRESQLKERLEKYODIHSIHILARKPNLGLKSCDGNQELLNFLRSLDIE--	VATRL	1070
Db	1541	RGIKEALQKIQK-HLDYITQACLKNK-DVAIIITLNENE--NOVTRDIVENWSVEEQA	1596	
Qy	1071	QKGGLGVYEETSEFEARVISLEKLKDFGECVIALOASVIKKFLOGFNAPKO-----	1121	
Db	1597	MENDLSVLQOVEDLERRVAS-----ASLQ-----	VKGMCMCPASEREDLVY	1638
Qy	1122	-----KRKLQSEDSAKTEEVDDEKKWVE--EAKVAS	1151	
Db	1639	FEHKSFTKLCKEHGDEFTGEDSSAHALERKSDNPLOJAVTRLADLERNIERIEEDIAP	1698	
Qy	1152	ALEKWKTAIREAQTFSRMHVLLGMDACIKWDMSAENARCVKCPKGDEDKLILCECNK	1211	
Db	1699	GLRYVRREALSEARSAAQCIOLOQSAIAWEKSIMKVYCQICRKGDNEBLLLDCGCDK	1758	
Qy	1212	AFHLFCRLPALYEPDGEWCPCACPATA-----RRNSGRGRNYTESASE-----	DS	1258
Db	1759	GCHYYCHRPKTIIPDGDNWFCPACIAKASGOTLKIKLHWGXKTNESKGGKVTLTGDT	1818	
Qy	1259	EDDESDBEEEEEEEBEEDYEVAGLRRLRPRTIRIGKHSVIPPAARGRRPGKKPHSTRR	1318	
Db	1819	EDEDS-----ASTSSLLKRGNKDLQKRR	1841	
Qy	1319	SOKPAPPVDAEVDLVLQTKRSSRRROSLELOKCEEILHKIVKYSVSFPREPVRTDEAE	1378	
Db	1842	MEENTIS--INLSKQESTSVKKPKRDDSKDALCSMLTETHEDAWPFLLPNLKLVP	1899	
Qy	1379	DYDVITHPMDFQTVQNKCSGYSRVSQEFLTMKQVFNTAEVYN	1423	
Db	1900	GKKVKIKKPMDFSTIREKLSGSGYPNETALDVLRFVDFNCETFN	1944	

RESULT 22

```

US-10-702-148-21
; Sequence 21, Application US/10702148
; Publication No. US20040063145A1
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042002
; CURRENT APPLICATION NUMBER: US/10702,148
; CURRENT FILING DATE: 2003-11-05
; PRIOR APPLICATION NUMBER: PRIOR APPLICATION NUMBER: US/09/839,479
; PRIOR FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 21
; LENGTH: 1972

```

Db 1021 LINEACSVSVSEIDKNDYNSNLRDKVV-----EGKRLKRIIHAKKTKRSTSGI 1076
QY 816 ENG-----LG-----KTRKKRIIVKFPQVPTAEADMISAVKSRRLAIAQAKREIOE 864
Db 1077 DLGEEQHPLGTTPGKRKRKGGSDYDDDDDDDD-----QGDDEDEE 1123
QY 865 --REMKVLERQAEERIRKHAABK-----AFQEGIAKAVLWRTPTGT 909
Db 1124 DXEDQKGKTDICEDEDEGDAASVEELEKQIEKLSKQSQYRRLKFDASHSLRSVMFGP 1183
QY 910 DRNHRYWLFSDVPGLFTE-----KGWH-----DSIDY 939
Db 1184 DRYRRYWL-PRCGIFVEGMESGEGLEEIAKERKLLKAESVQIKEMFETSGDSLNC 1242
QY 940 RENHHC-----KDHT--VSGDEDYCPRSK-----KANLGKNA--- 969
Db 1243 SNTDHCQEKEDLKKONTNLFQKPGSPKSLKLEVAKMPPESEVMTPKPNAGANGCTL 1302
QY 970 -----SMNTOHGDTATEVAVE-----TTTPKQOQ 992
Db 1303 SYQNSGKHSLGVSQSTATQSNVEKADNNLFTNGSSGPGKFSYPLPNDQLLTKLTEKNRQ 1362
QY 993 NLWFL-----CD-----999
Db 1363 --WFSLLPRTPCDDTSLTHADMSTASLVTPQSPPSKSPSPTPAPLGSQAQNPVGLNPPA 1420
QY 1000 -----999
Db 1421 LSPLOVKGVGMGLQFCGPTGVVTSNIPFTLSVPSLGLSGLSEGNNGNSPLTNSVASS 1480
QY 1000 -----SQKDELINCLHP 1013
Db 1481 KSESPVPQNEKATSAQPAAVEAVKVPDFSPKPIPEEMQFGWRIIDPEDLKALLKVLHL 1540
QY 1014 QGIRSQLKREKRYODIISHILARKPNLGLKSCDGNOLNLFRLSDLIB--VATRL 1070
Db 1541 RGIREKALQKQOK-HLDYITQACLKVK-DVAIIELNENE--NQVTRDIVNSVVEQA 1596
QY 1071 QKGGLGYVETSEFEARVISLEKLKDFGECVIALQASVIKKFLOGFMAPKQ-----1121
Db 1597 MEMDLSVLQOVEDLERRVAS-----ASLQ-----VKGMCEPASEREDLVY 1638
QY 1122 -----KRRKQSEDSAKTEVDEEKKWZ--FAKVAS 1151
Db 1639 FEHKGFTKCKHGDGEFTGEDESSAHALERKSDNPLDIAVTRLADLERNIERRIBEDTAP 1698
QY 1152 ALEKWKTAIREAQTSRMHVLMLDACIKWDMSAENARKVCYKPKGDDKLILCDECNK 1211
Db 1699 GLRVWRALSEARSAQAVALCIIQLOKSIANKVYQICRKGDNBEILLLCDGCK 1758
QY 1212 AFHLFCLRPALYEPDGEWQCPACOPATA-----RRNSRGNYTEESASE-----DS 1258
Db 1759 GCHTYCHRPKITIPDGMWFCFACIAKASGQTLKIKLHVKKTKNESKKGKVVLTGDT 1818
QY 1259 EDESEDEEBEEBEEDYEVAGLRLPRKTIRKHSVIPPAARSGRRPKKPHSTR 1318
Db 1819 EDEDS-----ASTSSSLKRGNKDLOKRAK 1841
QY 1319 SOPKAPPVDDAEVDELVLQTKRSSRQSLQKCEILLHKVYKRFSPFPVPTRDEAE 1378
Db 1842 MEENTS--INLSKQESFTSVKPKRDDSKDLALCSMLITEMETHEDAWPFLPVLNKLVP 1899
QY 1379 DYVDWTHPMDFQTQVONKSCGSYSRVQBFLLTDMKQVFTNAEVYN 1423
Db 1900 GYKVKIKKPMDFSTIREKLSGGYPNLETALDVLVDFDNCETFN 1944

RESULT 23
US-10-087-192-666
; Sequence 666, Application US/10087192
; Publication No. US20020182586A1
; GENERAL INFORMATION:

; APPLICANT: Morris, David W.
; APPLICANT: Engelhard, Eric K.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
; FILE OF INVENTION: CANCEER
; FILE REFERENCE: 529452000122
; CURRENT APPLICATION NUMBER: US/10/087,192
; CURRENT FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 2059
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 666
; LENGTH: 1873
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-087-192-666

Query Match 6.7%, Score 531.5; DB 13; Length 1873;
Best Local Similarity 19.1%; Pred. No. 4.4e-21;
Matches 261; Conservative 183; Mismatches 438; Indels 483; Gaps 40;

QY 415 OKSTGNSKSPKGLKTPKTKMQMTLLDMAKGTQ---KMTRAPRNSGGTPTSTSSPHKHL 471
Db 616 QAITGKRGRPR---NTEAKTKVEPKVKGGRPRPKVKITEL-----LNKTDNRPKXL 666
QY 472 PPAALHLIAYKENKNDREKRSALSVCISKTAAR-----LLSSEDRARLPBEL 518
Db 667 EA-----QETLNEEDKAK-----IAKSKKMRQKVQVQECOTTIQQOARNKRKQET 712
QY 519 RSLQVKRYELLEHKKRWASMSSEQRKEYLKKREELKKLKE---KAK-----563
Db 713 KSLQKQAKKSKKKEGKTQKQELKEKVKREKVKMKKEEVTAKKAPACKADKTLATQ 772
QY 564 --ERREKMLELEKQKRYEDQELTG-KNLPAFLVDTPTEQ--LPNTLFGVAMVVBFL 617
Db 773 RLREERQOQWILEMKKPTEDMCLTDHQPDPDSRV---PGLTLPGSFAFSDCLTIVEFL 829
QY 618 SCYSGLLPDAQYPIITAVS-LMEASADKGFYLNRLVILLQTLLODETAEDYG 676
Db 830 HSFQKVLGDFPAKDVPSLGVQLQEGLLCQGDLSLGEVQDLLVRLKALHDPGPPSYCQSLK 889
QY 677 ELGMKLSIPLTLHSVSELVRLCLRRSDVQSESGSDTDNDKDSAAAFEDNEVQDFLEKL 736
Db 890 ILGEKVSIEPLTRDNVSEILRCFLMAYG-----VEPALCDRL 926
QY 737 ETSEPFELTSBEKLIQILFALCHRI---LMTYSVQDHMETRQMSAELW--KERLAVLKE 790
Db 927 RTQFQAQPPQQAALVAFVLHVLNGSTLIIINEIDKTLIESMSSYRKWKWIVEGRLRLK- 985
QY 791 ENDKRAEKQKREKEMKENGKNGKVENGLKTKDRKRIKVKFEPQVDTBAEDMISAVKGR 850
Db 986 ---TVLAKRTKGRSEVMEGP-----EECLGRRRSRIMEETSGMEEBEEESIAAVPGR 1037
QY 851 LLAIOAKKEREIOEREMKV-KLROAEEERIRKHAABEKAFOEGIAKAKVMRPTPTGT 909
Db 1038 -----GRDGEVDATPASSIPELERQIEKLSKR-----QLFFRKLLHSSQMLRAVSLGQ 1086
QY 910 DRNHRYWLFSDVPGLFTE-----KGWHDSIDYRF-----941
Db 1087 DRYRRYWL-PYLAGI FVEGTGNLVEEVIKKTDSLKVAHAASLPAFLFSMKMELAG 1145
QY 942 -----NHHCKDHTVSGDEDYCPRSK-----961
Db 1146 SNTTASPARARGRPRTKPGSMQPRHLKS-PVRQDSEQQAQLQPEAQLHAPAQPOFQ 1204
QY 962 -----KANLGKN-----ASMTQHGCTATEVAVETTTPKQG- 991
Db 1205 LQLOLQSHKGFLEQSGSPLSIGSOHDLQSASFLSWLSQTSOHSLLSSSVLTPDSSPK 1264
QY 992 -----QNLWFLCDSQ-----1001

Db 1265 LDPAPSOPEPEDEABESSDPQALWFNISAMPCNAAPTTPPAVSEDOPTSPQOLAS 1324
Qy 1002 ----- 1001
Db 1325 SKPMNRPSAANPCSPVQFSSTPLAGLAPKRRAGDGPMPQSPGGLQPKRGRPPSKFFK 1384
Qy 1002 ----- KELDNLNCLHPQOIRISQSLKRLKRYQDI 1032
Db 1385 QMBQRYLTQTAQFVPPMCGMWWIRDPMLDAMLKALHPRGIREKALHKLHNK-HRDF 1443
Qy 1033 IHSIHL--ARKPNLGLKSCDGNOR-LLNFLRSLDIEVATRLQKGLGYVEETSEFARVI 1089
Db 1444 LQEVCLRPSADPIFEPRQLPAFOEGIMSPKE-----KTYETDLAVLQWVEELQRVI 1497
Qy 1090 S-----LEKLKDFGECV----- 1101
Db 1498 MSDLIQIRGWTCPSPDSTREDLAYCEHLSDSQEDI TWGRGREG LAPQRKTNPDLAVMR 1557
Qy 1102 -IALQASVIKFFLOGMAPKQKRKLQSDSASATEEVDKQWVE-----EA 1147
Db 1558 LAALEQNVERRYLRPLWP-----THEVWLEKALLSTPNGAPEGTTTTEISY 1603
Qy 1148 KYASALEKWKTAIREAQTSRMHVLLGMLDACIKWMSAENARCKYCPKKGEDDKLILCD 1207
Db 1604 EITPRIRVWRQTLERCRSAQVCLGLQGLERSIAWKSVMKVTCLYCRKGDNDDEFLLCD 1663
Qy 1208 ECKNAFLCLRPALYEVDPGEWQCPACOPATARRNSRGRNYTEESASEDESDDEE 1267
Db 1664 GCDRGCHIVRPMKMAVEGDFCTVC-----LAQVGEFTQKPG 1705
Qy 1268 EEEEEEEEDYEV---AGLRLPRKTIIRGKHSVIPPAARSGRRPKKPHSTRSQPKAP 1324
Db 1706 FPKRGQKRKSGVSLNFSGDRRRVLLRGRES--PAAGPRYSEGLSPSKRR----- 1757
Qy 1325 PVDVDAEVDLVTQKSSRRQSLELOKCEILHKIVKYRFSWPPRPVTRDAEDYDVI 1384
Db 1758 -----LSMRNHHSDLTFCEIILMEMESHDAAPFLEPVNRLVSGYRRII 1802
Qy 1385 THPMDFTQVONKSCGSYRSVQBFLLDMKQVFTNAEVYNGRSHV 1429
Db 1803 KNPMDFSTWRLLRGYTSSEEFADALLVFNDCQTFNEDDSEV 1847

RESULT 24

US-09-839-479-70
; Sequence 70, Application US/09839479
; Publication No. US20020039779A1
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042002
; CURRENT APPLICATION NUMBER: US/09/839,479
; CURRENT FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/Jp98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 70
; LENGTH: 1876
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-839-479-70

Query Match 6.5%; Score 518.5; DB 9; Length 1876;
Best Local Similarity 19.4%; Pred. No. 2.4e-20;
Matches 267; Conservative 177; Mismatches 429; Indels 501; Gaps 44;

Qy 415 QKSTGNSKSPKGLKTPKTKMQMTLLDMAKGTQ---KMTFAPRNSGGTPTRTSSKPKHKL 471
Db 619 QAITGRGRPR---NTEKAKTKEVPKVRGRPPKVKITEL-----LNKTDNRPLKKL 669
Qy 472 PPAALHLIAYIKENKDRKRSALSCVISKTARLLSSSEDRARLPBELRSIVQRYELLEH 531
Db 670 EA-----QETLNEEDKAKIAKSKKQVRQVQGECLTTTIGQARNKFQKETSLSKH 720
Qy 532 KKRWASMSBEQKEYLKKREELKKLKEKAKERREKE-----M 570
Db 721 KE--AKKKSABERCKTQEKLEKVKREKKEKVKKEEVTAKPACADKTLATQRR 778
Qy 571 LERLEKQ-----KRYEDQBLTG-KNLPAPRLVDTPGEG--LPNTLFGDVAMVVFSLC 619
Db 779 LEERQKQMILEBMKKPTEDMCLTDHQLPDFSrv---PGLTLPSCGAFSDCLTIVEFLHS 835
Qy 620 YSGLLLPDAQYPTITAVSLMEALSADKGGFLYNRLVLILOTLQTLLODEIAEDYGE-- 677
Db 836 PKGVLFDPKAOVPSLGVLOEGLLCQDLSL---GEVQDLLVRLKKAALHDPGFPSPYCSQK 892
Qy 678 --LGMKLSIPLTLHSVSELVRLCLRRSDVQBESEGSDDDNKDSAAFEDVEQDFLEK 735
Db 893 KILGEKVSIPILTRDNVSEILRCFLMAYG-----VEPALCDR 929
Qy 736 LETSEFFELTSBEKLQILTALCHRI---LMTYSVQDHMETRQMSAELW--KERLAVLK 789
Db 930 LATQFPQAPQPOQKAVALFAPVHELNGSTLIINEIDKTLESMSYRKNKWIVEGRLLRLK 989
Qy 790 EENDKGRARQKQKWEAKKNKENGKVENGLGKTDRKRIVKPEQVDTEADMISAVKSR 849
Db 990 ---TVLAKRTGRSEVEM---GRPECLGRRRSRIMEETSGMEEEEEESIAAVPGR 1040
Qy 850 RLLATQAKKERIIOEREMKV-KLERQAEERIRKHKAAAKAFAOEGIAKAKLVMRTPIG 908
Db 1041 R-----GRDGEVDATASIPLELRQIEKLSR-----QLFRKKLLHSSQMLRAVSLG 1089
Qy 909 TDRNHNRYW-----LFSDEV-----PGLFTIEK----- 930
Db 1090 QDRYRRRYVWLPYLAGIFVEGTEGNLVPEVICKETDSLKVAHAASLNPALEFMSKMLAG 1149
Qy 931 ----- 930
Db 1150 SNTTASSPARASRPLKTKPGFMQPREPKSPVRGODSEQPQALQPEAQLHVPAQPQQL 1209
Qy 931 -----GWV-----HDSI----- 937
Db 1210 QLQLOSHKGFLEQSGSLSGOSQHDLSQSAFLSWLSQTHSSLLSSSVLTDPDSSPGKLD 1269
Qy 938 -----DYRFNHHCK-----DHTVSGDEYCYCRSKKANLG 966
Db 1270 PAPSQPPEPEPEDEABESSPDLQAFWFNISAQPCNAAPTTPPLAVSEDO---PTPSPQOLA 1326
Qy 967 KNASWN----- 972
Db 1327 SSKPMNRPSAANPCSPVQFSSTPLAGLAPKRRAGDGPMPQSPGGLQPKRGRPPSKFF 1386
Qy 973 ---TOHGTATEVAVETTPKQGNLWFLCDSQKSELNCLHPQOIRISQSLKRLKRYQ 1030
Db 1387 KOMEQORVLTQTAQVPPMCGMWWIRDPDEM-LDAMLKALHPRGIREKALHKLHNK-HR 1444
Qy 1031 DIHSIHLKAPENLGLKSCDGNQELLNFLRSLDIEVATRL---QKG-----G 1074
Db 1445 DFLQEVCL--RPS-----ADPIPEPRQLPAFOEGIMSPKPKTYETD 1485
Qy 1075 LGYVEETSEFEARVIS-----LEKLKDFGECV----- 1101
Db 1486 LAVLQWVELEQIRVIMSDIQRGWTCPSPDSTREDLAYCEHLSDSQEDITWRGPGREGLA 1545
Qy 1102 -----IALQASVIKFFLOGMAPKQ-----KRRKQSEDA---KTEEVD 1138
Db 1546 PORKTTNPDLAVMLALAEQNVKRYLRPLWPTHEWVLEKALLSTPNGAPEGTTTTEIS 1605
Qy 1139 BEKQWVEEAKVASALEKWKTAIREAQTSRMHVLLGMLDACIKWMSAENARCKVCPKKG 1198

Db 1606 YE-----ITPRIRIWRQTLQRCBSAAHVCLGHLERSIAWEKSVNKTCLVCKGD 1657
QY 1199 EDDKILDCENKAFHLCRLPALYVDPGEWQCPACQATARRNSRGRNYYTESASEDS 1258
Db 1658 NDEFLLLCGDGRGCHYCHRPKMEAVPBGDFCTVC-----LAQOV 1699
QY 1259 EDDSEDEEEEBEEDYEV---AGLRPRKTIIRKHSVIPPAASRRRPGKPKHS 1315
Db 1700 EGEFTQKPGFPKRGKRGKSGYSLNFSBGDRRRVLLKGRES---PAA-----GPRYSE 1750
QY 1316 TRRSQPKAPVDVAEDELVLQTKSSRRQSLEQCEILHKIVKRYFSPFRBPVTRD 1375
Db 1751 ERLSPSKRPL-----SMRHHSDLTFCCEIILMEMESHDAWFFLEPVNPR 1796
QY 1376 EAEDYDVITHPMDFTQVQNKSCGYSYRSVQBFLLTDMKQVFTNAEYVNCGRSHV 1429
Db 1797 LVSGYRRIIKNPDMFSTMRERLLRGYTSSEFAADALLVFNCTFNEDDSEV 1850

RESULT 25
US-10-376-537-71
; Sequence 71, Application US/10376537
; Publication No. US20030224405A1
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042001
; CURRENT APPLICATION NUMBER: US/10/376,537
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US/09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 71
; LENGTH: 1876
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-376-537-71

Query Match 6.5%; Score 518.5; DB 15; Length 1876;
Best Local Similarity 19.4%; Pred. No. 2.4e-20;
Matches 267; Conservative 177; Mismatches 429; Indels 501; Gaps 44;

QY 415 QKSTGNSKSPKGLTKPTKMKQMTLLDMAKGTQ---KMTAPRNSGGTPTRTSSKPHKL 471
Db 619 QAITKGRGRPR---NTEAKTKEVPKVRGRPRPKVITEL-----LNKTDNRPLKKL 669
QY 472 PPAALHLIAYKENDKRKSALSVISKARTALLSSDRALPELRLSLVKRYELLEH 531
Db 670 EA-----QETLNEDKAKIAKSKKQKQKQVQGECLTTIOGARNKFKQETKSLKH 720
QY 532 KKRWASMSBEQRYLKKRRELLKKLKEKAKERREKE-----M 570
Db 721 KE--AKKSKAEKRGKTKQELKEKVKEKEKVKEEYVTKAPACKADKTLATQRR 778
QY 571 LERLEKQ-----KRYEDQELTG--KNLPAPRLVDTPEG--LPNTLFGDVAMVVEFLSC 619
Db 779 LEEROKQOQILBEMKKPTEDMCLTDHQLPDPFSRV---PGLTLPAGFSDCLTIVEFLHS 835
QY 620 YSGLLLPDAQYIPITAVSLMEALSADKGGFLYNRLVILLOTLLQTLLODETAEDYGE-- 677
Db 836 PKGVGFDPADKDVPSLGVVQEGLLCQGDLSL---GEVQDILLVRLKAAHLDPGFPSYCSK 892
QY 678 --LGMKLSPIPLTHSVSELVRLCLRRSDVQBESEGSDDTDNKSAAAFEDNEVDQEFLEK 735
Db 893 KILGEKVSEIPLTRDNVSEILACFLMAY-----VEPALCDR 929

RESULT 26
US-10-702-148-70
; Sequence 70, Application US/10702148

QY 736 LETSEFFELTSEEKLIILTALCHRI-----LMTYSVQDHMETRQOQMSABELW--KERLAVLK 789
Db 930 LRTQFQAPQPPQQAALVAPVPHELNGSTLIINEIDKLTLESMSYRKNKVIIEGGLRLUK 989
QY 790 EENDKRAEKQKREMEAKNKENGKVENGLGKTDKKGKIVKFEPOVDTEAEDMISAVKSR 849
Db 990 ----TVLAKRTGRSEVEM-----GRPECLGRRSSRRIMEETSGMEEEEEESIAAIVPR 1040
QY 850 RLLAIQAKKEREIOEREMKV--KLERQABEERIIRKHAAAFAQFQEGIAKAKLVMRTTIG 908
Db 1041 R-----GRDGEVDATASSIPELERQIEKLSKR-----QLFFRKLLHSSQMLRAVSLG 1089
QY 909 TDRNNRYW-----LFSDEV-----PGLPIEK----- 930
Db 1090 QDRYRRYVWLPYLAGIFVEGTGNLVPEEVIKKETDSLKVAHAASLNPAALFPMKVELAG 1149
QY 931 ----- 930
Db 1150 SNTTASSPARARSRLKTKPGFMQPREFKSPVRGQDSEQQAQLOPEAQLHVPAQPQL 1209
QY 931 -----GVV-----HDSI----- 937
Db 1210 QLQLQSHKGFLEQEGPLSLGSOQHDLSQSAFLSWLSQTSQHSLLSSSVLTDPDSSPGKLD 1269
QY 938 -----DYRFNHCK-----DHTVSGDEDYCPRSKKANLG 966
Db 1270 PAPSOPPEEPFDEABSSPDLQAQFWFNISAQMPCNAAPTPLAVSEDO---PTSPQQOLA 1326
QY 967 KNASMN----- 972
Db 1327 SSKPMNRPSAANPCSPVQFSSTPLAGLAPKRAGDPGEMPSQPTGLQPKRGRPPSPKF 1386
QY 973 --TQHGTAIVAVETTPKQGNLWFLCDOSKELDELNCLHPQGIRESQLEKREKYQ 1030
Db 1387 KQMEQVLTQLTAQVPVPEMCGMWIIPDPM-LDAMLKALHPRGIREKALHKLHNK-HR 1444
QY 1031 DIHSIHILARKPNLGLKSCDGNQELLNFRSLDLEVATRL---QKG-----G 1074
Db 1445 DFLQEVCL--RPS-----ADPIPEPRLPAFOEGINWSWSPKETYETD 1485
QY 1075 LGYVEETSEFEARVIS-----LEKLKDFGECV----- 1101
Db 1486 LAVLQWVEELQQRVIMSDLIQIRGWTCPSPDSTREDLAYCEHLSDSQEDITWRGREGLA 1545
QY 1102 -----IALQASVIKKFLQGFMAPKQ-----KRRKLQSEDSA---KTEEVD 1138
Db 1546 POKTTNPLDLAVMRLAALAEQNVKRRYLREPLWTHMWLEKALISTENGAPEGTTTIS 1605
QY 1139 BEKQWVEAKVASALEKWKTAIREAQTFSRMHVLLGMLDACIKWDMSAENARCKVCPKG 1198
Db 1606 YE-----ITPRIRIWRQTLQRCBSAAHVCLGHLERSIAWEKSVNKTCLVCKGD 1657
QY 1199 EDDKILDCENKAFHLCRLPALYVDPGEWQCPACQATARRNSRGRNYYTESASEDS 1258
Db 1658 NDEFLLLCGDGRGCHYCHRPKMEAVPBGDFCTVC-----LAQOV 1699
QY 1259 EDDSEDEEEEBEEDYEV---AGLRPRKTIIRKHSVIPPAASRRRPGKPKHS 1315
Db 1700 EGEFTQKPGFPKRGKRGKSGYSLNFSBGDRRRVLLKGRES---PAA-----GPRYSE 1750
QY 1316 TRRSQPKAPVDVAEDELVLQTKSSRRQSLEQCEILHKIVKRYFSPFRBPVTRD 1375
Db 1751 ERLSPSKRPL-----SMRHHSDLTFCCEIILMEMESHDAWFFLEPVNPR 1796
QY 1376 EAEDYDVITHPMDFTQVQNKSCGYSYRSVQBFLLTDMKQVFTNAEYVNCGRSHV 1429
Db 1797 LVSGYRRIIKNPDMFSTMRERLLRGYTSSEFAADALLVFNCTFNEDDSEV 1850

```
; Publication No. US20040063145A1
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTONAL REGULATOR
; FILE REFERENCE: 06501-042002
; CURRENT APPLICATION NUMBER: US/10/702,148
; CURRENT FILING DATE: 2003-11-05
; PRIOR APPLICATION NUMBER: PRIOR APPLICATION NUMBER: US/09/839,479
; PRIOR FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 09/418,710
; PRIOR FILING DATE: 1998-04-17
; PRIOR FILING DATE: 1999-10-15
; PRIOR FILING DATE: 1997-10-24
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 70
; LENGTH: 1876
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-702-148-70

Query Match      6.5%; Score 518.5; DB 15; Length 1876;
Best Local Similarity 19.4%; Pred. No. 2.4e-20;
Matches 267; Conservative 177; Mismatches 429; Indels 501; Gaps 44;

QY 415 QKSTGNSKSPKGLKTPKTKOMKMTLLDMAKGTQ---KMTAPRNSGGTPTRTSSKPHKHL 471
DB 619 QAITGKRGPR---NTEKATKEVPKVGKRGPRPKVKITEL-----LNKTDNRPLKKL 669
QY 472 PPAALHLIAYKKNKDKRSGSALSVCISKTALLSSDRAPLPELSLVQKRYELLEH 531
DB 670 EA-----QETLNEEDKAKTAKSKKQMRQVQGECLTTIQGARNKFKQETKSLKH 720
QY 532 KKGWASMSBEQRKYLKKREELKKLKKKAKERREKE-----M 570
DB 721 KE--AKKSKAEKERGKTQKELKEKVKREKKEKEVEVTKAKPACKADKTLATQRR 778
QY 571 LERLEKQ-----KRYEDQELTG--KNLPAPFLVDTPEG--LPNTLFGDVAMVVEFLSC 619
DB 779 LEERQKQOILEEMKKEPTDMCLTDHQLPDPFSRV---PGLTLPFGAFSDCLTIVEFLHS 835
QY 620 YSGLLLPDAQYPITAVSLMEALSADKGGFLYLNRLVILLQTLLOLDEIAEDYGE-- 677
DB 836 PGKVLGFDPAKOVPSLGVJQEGLLCGDLSL---GEVQDILLVLLKAAHLDPGPPSYCQSK 892
QY 678 --LGMKLSIEPLTLHVSVELVRLCLRRSDVQEESEGSTDDNKKDSAAFEDNEVQDEFLEK 735
DB 893 KILGEKVSEIPLTRDNVSEILRCFLMAYG-----VEPALCDR 929
QY 736 LETSEFETLTSEKQLQLTALCHRI---LMTYVSQDHMETROQMSAELW--KERLAVLK 789
DB 930 LRTQPPQAQPPQKAAVLAFFVHELNGSTLIINEIDKTLSSMSYRNKNKWIVEGRULRLK 989
QY 790 EENDKKRAEKQKKEKAKKENGKENGKVLGKTDKRRIVKPEPQVDTAEDMISAVKSR 849
DB 990 ----TVLAKRTGSEVEM-----GRPEELGRRRSRIIMEETSGMBEEEEESIAAVPR 1040
QY 850 RLIAIOAKKEREQEREMKV--KLEOAAEERIRKHAAKAAKAFQEGIAKAKLVMRTPTIG 908
DB 1041 R-----GRDGEVDATASSIPELERQIEKLSK-----QLFPRKLLHSSQMLRAVSLG 1089
QY 909 TDRNHNRYW-----LFSDEV-----PGLFTEK----- 930
DB 1090 QDRYRRRYVWLYPYLAGIFVEGTGNLVPEVIEKTDLSLKAHAHSLNPALEFSMKMLAG 1149
QY 931 ----- 930
DB 1150 SNTTASSPARARRPLTKPGFMQPREFKSPVRGQDSQFOAQLQLOPEAQLHVPAPQPQL 1209
```

```
QY 931 -----GWV-----HDSI----- 937
DB 1210 QLQLQSHKGFLEQEGSPLSLGQSQHDLSQSAFLSWLSQTHSSLSSTSSSVLTDPDSSPGKLD 1369
QY 938 -----DYRNHHCK-----DHTVSGDEDYCPRSKKNILG 966
DB 1270 PAPSQPEEPPEDEABESSPDLOAFWNFISAQPCNAAPTPLAVSBDQ---PTPSPQOLA 1326
QY 967 KNASNN----- 972
DB 1327 SSKPMNRPSAANPCSPVQSSSTPLAGLAPKRRAGDPGEMPQSTGLGQPKRRGRPPSKFF 1386
QY 973 ---TQGTATEVAVETTPPKQGNLWFLCDSQKELDELNCLHPQGIRESQKLERLEKRYQ 1030
DB 1387 KQMEQRLVLTQTAQPVPPPEMCGMMWIPDPFM-LDAMLKALHPRGIREKALHKLK-HR 1444
QY 1031 DIIHSIHLARKPNLGLKSCDGNQELLNFLRSLIEVATRL---OKG-----G 1074
DB 1445 DFLQEVCL--RPS-----ADPIPEPQLPAPQEGIMSWSPKETYETD 1485
QY 1075 LGVYEETSEFEARVIS-----LEKLDKDFGECV----- 1101
DB 1486 LAVLQWVELEQVRVMSDLQIRGWTCPSPDSTREDLAYCEHLSDSQEDITWRGPGREGLA 1545
QY 1102 -----IALQASVIKKFLQGFMAKQ-----KRRKLOSEDSA---KTEEVD 1138
DB 1546 PQKTTNPDLAVMLAALAEQNVKRYLREPLWPTHEWVLEKALLSTPNGAPBGTTEIS 1605
QY 1139 BEKKWVEAKVASALEKWKTAIREAQTFSRMHVLLGMLDACIKWMSAENARCKVCPKKG 1198
DB 1606 YE-----ITPRINWRQTLQRCRSAHVCLCLGHLERSIAWEKSVNKTCLVCRKGD 1657
QY 1199 EDDKLILCDCKNAHFLCLRPALYEPDGEWOCQACQATARRNSRGRNYTEESASEDS 1258
DB 1658 NDEFLLLCDGCRGCHYCHRPMEAVPEGDFCTVC-----LAQOV 1699
QY 1259 EDESDSEEEEEEEEDYEV---AGLRLPRKTIRGKHSVIPPAAISGRRPGKKPHS 1315
DB 1700 EGEEQTKPGFPKRGKRGKSGYSLNFSSEGDRRRVLLKGRES---PAA-----GPRYSE 1750
QY 1316 TRSQPKAPPVDDAEVDELVLQTKSSRRQSLEQKCEILHKIVKYRFSWPPREPVTRD 1375
DB 1751 ERLSPSKRRPL-----SMRNHSDLTFCETILMEMESHDAWPPLEPVNPR 1796
QY 1376 EAEDYVDVITHPMDFQTVONKSCSGSYRSVQBFELTDMKOVFTNAEYVNCRGSHV 1429
DB 1797 LVSGTRRIKPNMDFSTMRERLLRGYTSSEEFADALLVFDNCQTFNEDDSEV 1850

RESULT 27
US-09-839-479-13
; Sequence 13, Application US/09839479
; Publication No. US20020039779A1
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTONAL REGULATOR
; FILE REFERENCE: 06501-042002
; CURRENT APPLICATION NUMBER: US/09/839,479
; CURRENT FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 1876
; TYPE: PRT
; ORGANISM: Homo sapiens
```

```

; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(1878)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-09-839-479-13

Query Match      6.4%; Score 512.5; DB 9; Length 1878;
Best Local Similarity 19.3%; Pred. No. 5.1e-20;
Matches 263; Conservative 179; Mismatches 443; Indels 479; Gaps 42;

QY 415 QKSTGNSKSPKKGLTKPTKQMTLLDMAKGTQ---KMTAPRNSGGTPTRTSSPKHKL 471
DB 619 QAITGKGRPR---NTEKAKTEVPKVKGRGPRPKVKITEL-----LNKTDNPLKKL 669
QY 472 PPAALHLIAYKENKDRKRSALSCVISTARLLSSEDRARLPEELSLVQKRYELLEH 531
DB 670 EA-----QETLNEEDKAKIAKSKKMRQKVGEGCLTTIQGAQNRKQETKSLKH 720
QY 532 KKRWASMSEQRKEYLKKREELKKLKEKAKER---REKE-----569
DB 721 KE--AKKKSABEKKGKTKQELKKEKVKREKKEKVKMKKEKEVTKAKPACKADKTLATQR 778
QY 570 MLERLEKQ-----KRYEDQELTG-KNLPARFLVDTPEG--LPNTLFGDVAMVVEFLS 618
DB 779 RLEERQKQOMILEEMKKPTEDNCLTDHQPLPDFSRV---PGLTLPGAFSDCLTIVEFLH 835
QY 619 CYSGLLLPDAQIPITAVS--LMEALSADGGGFLYNRVLVILLQTLLODEIAEDYGE 677
DB 836 SFGKVLGFDPAKDPVPSLGLVQSGLLCQDGLGEVDLLVRLKAAALHDPGPPSYCQSLKI 895
QY 678 LQMKLSEIPLTLHSLVSRVLCRLRSRDVQEESEGSTDDNDKDSAAFEDNEVQDEFLEKLE 737
DB 896 LGEKVEIPLTDNDNSEILURCFIMAYG-----VXPALCDDRUR 932
QY 738 TGEFFELTSEEKLQILITLCHRI---LMTYSVQDHMETROQMSAELW--KERLAVLKEE 791
DB 933 TQPFQAQPPQKAAVLAPPVHELNGSTLIINEIDKTLSEMSYGRKNKWIVEGLRLAK-- 990
QY 792 NDKKAAEQKRVKEMAKNKENGVNGLGKTRKRIIVKFEPOVTEAEDMISAVKSRRL 851
DB 991 --TVLAKTRGRSEVM-----GRPECLGRRRSSRIMEETSGWEEBEEESIAAVPGR-- 1042
QY 852 LAIQAKKEREIQEREMKV-KLRQAEERIRKHAAAKAFQEGIAKAKLVMRRTPIGTD 910
DB 1043 ---GRRDEVDATASSPELERQIEKLSK-----QLFFRKLHLSSQMLRAVSLGD 1092
QY 911 RNHNRYWLFSDVPGLFIB-----KGWVHDSID-----938
DB 1093 RYRRRYWVL-PYLAGIFVETGEGNLVPERVIKKTDSLKVAHAASINPALFSMKMELAGS 1151
QY 939 -----YRFNHCKDHTVSGDEDYCPRSK-----961
DB 1152 NTTASSPARARSRLTKTGFQMPRHFKS-PVRGQDSEQPAQLOPEAQLHVPAQFPQOL 1210
QY 962 -----KANLGN-----ASMTQHTGATEVAVETTPKQG-- 991
DB 1211 QLQLOSHKGFLQEGSPLSGQSHDLSQSASFLSMLSTQSHSLSSSVLTPDSSPGKL 1270
QY 992 -----QNLWFLCDSQ-----1001
DB 1271 DPAPQPPPEPDEAESSPDIAQAFWNISAQMPCNAAPTPTPLAVSEDOPTSPQOLASS 1330
QY 1002 -----1001
DB 1331 KPMNRPSAANPCSPVQFSSTPLAGLAPKRRAGDGPMPQSPGTGLGQPKRGRGPPSKFFKQ 1390
QY 1002 -----KEDELNCLNHPQIGRESQLEKLEKYQDII 1033
DB 1391 MQQRYLTQTAQVPPEMCSGWWIIPDDPEMLDAMUKALHPRGIRGKALHKLHUNK--HRDFL 1449
QY 1034 HSIHL--ARKPNLGLKSCDGNQE-LINFLRSLLIEVATRLQKGGYGYVEETSEFEARVIS 1090
DB 1450 QEVCLRPSADPIFEPRQLPAFQEGIMSPKE-----KTYETDLAVLQWVEELQQRVIM 1503
```

RESULT 28

```

US-10-376-537-13
; Sequence 13, Application US/10376537
; Publication No. US20030224405A1
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042001
; CURRENT APPLICATION NUMBER: US/10/376,537
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US/09/418,710
; FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 1878
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(1878)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-10-376-537-13

Query Match      6.4%; Score 512.5; DB 15; Length 1878;
Best Local Similarity 19.3%; Pred. No. 5.1e-20;
Matches 263; Conservative 179; Mismatches 443; Indels 479; Gaps 42;

QY 415 QKSTGNSKSPKKGLTKPTKQMTLLDMAKGTQ---KMTAPRNSGGTPTRTSSPKHKL 471
DB 619 QAITGKGRPR---NTEKAKTEVPKVKGRGPRPKVKITEL-----LNKTDNPLKKL 669
QY 472 PPAALHLIAYKENKDRKRSALSCVISTARLLSSEDRARLPEELSLVQKRYELLEH 531
DB 670 EA-----QETLNEEDKAKIAKSKKMRQKVGEGCLTTIQGAQNRKQETKSLKH 720
QY 532 KKRWASMSEQRKEYLKKREELKKLKEKAKER---REKE-----569
```

Db 721 KE--AKKSAEKEGKTKQEKLEKVKREKEKVKMKKESEVTKAKPACKADKTLATOR 778
QY 570 MLERLEKO-----KRYEDQELTG-KNLPAPRLVDTPG--LPNTLFGDVAMVVEFLS 618
Db 779 RLEERQKQOMILEEMKPTEDMCLTDHQLPDSRV---PGLTSPGAFSDCLTIVEFLH 835
QY 619 CYSGLLLPDAQYPITAVS-LMEALSADKGGFLYLNRLVILQTLQTLQDEIAEDYGE 677
Db 836 SFGKVLGFPDPAKDVPSGLVQEGLLCQGDGLGEVQDILLVRLKAAALHDPGFPSPQSLKI 895
QY 678 LGMKLSIELPLTLHVSUSELRLCLRRSDVOEESGSDTDNDKDSAAFEDNEVQDEFLEKLE 737
Db 896 LGEKVSIEPLTRDNVSEILRCFLMAYG-----VXPALCDRLR 932
QY 738 TSEFFELTSEELQIILTALCHRI---LMTYSVQDHMETRQOMSABLW--KERLAVLKEE 791
Db 933 TOPFOQPPQQAVALAPFVHELNGSTLIINEIDKTLSEMSYRKNKWIVEGLRLRLK-- 990
QY 792 NDKRAEKQKRMKEMAKKENGKVENGLGKTRKRIKIVKPEPQVDTAEADMISAVKSRRL 851
Db 991 --TVLAKTRGRSEVEM-----GRPECLGRRSSRIMEETSGMEEBEEESIAAVPGR- 1042
QY 852 LAIQAKKEREIOEREMKV-KLROAEERIRKHAAAEKAFQEGIAKAKLVMRPTIGTD 910
Db 1043 ----GRDGEVDATASSIPELERQIEKLSK-----QLFRRKLLHSSQMLRAVSLGOD 1092
QY 911 RNHNRYWLPFSDEVPGLFIE-----KGWVHDSID----- 938
Db 1093 RYRRRIWVL-PYLAGIFVGTGTEGVLNVPBEVIKETDSLKVAHAHASUNPALFSKMKMELAGS 1151
QY 939 ----YRFNHCKDHTVSGDEDYCPRSK----- 961
Db 1152 NTTASSPARARSPLKTKGFMQPRHFKS-PVRGQDSEQOQAQLQPEAQLHVPAPQPQL 1210
QY 962 ----KANLGKN-----ASMTQHGCTATEVAVETTPKQG-- 991
Db 1211 QLQLOSHKGFLQEGSPLSGQSQHDLSQSAFLMSLQTSQSHSLSSSVLTPDSSPGKL 1270
QY 992 ----QNLWFLCDSQ----- 1001
Db 1271 DPAPSQPPEPEDEAESPDLOAFWFNISAQMPCNAAPTPPLAVSEDOPTSPPOOLASS 1330
QY 1002 ---- 1001
Db 1331 KPMNRPSAANPCSPVQFSTPLAGLAPKRAGDPGEMPOSPTGLGQPKRRGRPPSKFFKQ 1390
QY 1002 ----KELDELLNCLHPQIGIRESQLEKREKRYQDII 1033
Db 1391 MEQRYLTQLTQAQVPPPEMCGWWWIPDPBMLDAMLKALHPRGIREKALHKHLNK-HRDPL 1449
QY 1034 HSHIL--ARKPNLGLKSCDGNQ--LLNPLRSDLIEVATRLQKGGLYVEETSEFEARVIS 1090
Db 1450 QEVCLURPSADPIFEPRQLPAFQEGINMSWSPKE-----KTYETDLAVLQVMBELESORVIM 1503
QY 1091 ----LEKLKDFCEV----- 1101
Db 1504 SDLIQIRGWTCPSPDSTREDLAYCEHLSDSQEDITWRGPGREGLAPORKTNPDLDAVMPL 1563
QY 1102 IALQASVIKFLQGFMAQKQKRLQSEDSAKTEEVDEEKKMVE-----EAK 1148
Db 1564 AALEQNVKRYLREPLWP-----THEVVLEKALLSTPNGAPEGTTTISYE 1609
QY 1149 VASALEKKWTAIREAQTFSRMHVILGMLDACIKWDSNAENACKVCPKKGEDDKLILCDE 1208
Db 1610 ITRIRIWRQTLQRCRCAHAHVCILGLHLERSIAWEKSVNKTCLVCRKGDNDFFLLCDG 1669
QY 1209 CNKAFLFLCLRALYVPPGQWQCPACQAPATARNRGRNYTEESASEDSDESEDEESE 1268
Db 1670 CDRGCHYIHRPXMVAEVPDGMFCTVC-----LAQQVEGEFTQKPGF 1711
QY 1269 EEEEEEEEDYEV---AGLRLRPRKTIIRGKHVIPPAAASGRRRPGKKPHSTRRSQKAPP 1325

RESULT 29

US-10-702-148-13
; Sequence 13, Application US/10702148
; Publication No. US20040063145A1
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042002
; CURRENT APPLICATION NUMBER: US/10/702,148
; CURRENT FILING DATE: 2003-11-05
; PRIOR APPLICATION NUMBER: PRIOR APPLICATION NUMBER: US/09/839,479
; PRIOR FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 1878
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(1878)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-10-702-148-13

Query Match 6.4%; Score 512.5; DB 15; Length 1878;
Best Local Similarity 19.3%; Pred. No. 5.1e-20;
Matches 263; Conservative 179; Mismatches 443; Indels 479; Gaps 42;

QY 415 QKSTGNSKSPKGLTKPTKMKQMTLLDMAKGTQ---KMTAPRNSGGTPTRTSSPKHL 471
Db 619 QAITGKRGRPR---NTEAKTKVVKRGRGRPPKVKITEL-----LNKTDNRPLKKL 669
QY 472 PPAALHLIATYKKNKDRDKRSALSCVISKTARLLSSDRARLPEELSLVOKRYELLEH 531
Db 670 EA-----QETNEEDKAKIAKSKKQKQVORGECLTTIQGAQNRKRQETHKSLKH 720
QY 532 KKRWSMSEEQRYKLLKKREELKKLKEKAKER---REKE----- 569
Db 721 KE--AKKSAEKEGKTKQEKLEKVKREKEKVKMKKESEVTKAKPACKADKTLATOR 778
QY 570 MLERLEKO-----KRYEDQELTG-KNLPAPRLVDTPG--LPNTLFGDVAMVVEFLS 618
Db 779 RLEERQKQOMILEEMKPTEDMCLTDHQLPDSRV---PGLTSPGAFSDCLTIVEFLH 835
QY 619 CYSGLLLPDAQYPITAVS-LMEALSADKGGFLYLNRLVILQTLQTLQDEIAEDYGE 677
Db 836 SFGKVLGFPDPAKDVPSGLVQEGLLCQGDGLGEVQDILLVRLKAAALHDPGFPSPQSLKI 895
QY 678 LGMKLSIELPLTLHVSUSELRLCLRRSDVOEESGSDTDNDKDSAAFEDNEVQDEFLEKLE 737
Db 896 LGEKVSIEPLTRDNVSEILRCFLMAYG-----VXPALCDRLR 932
QY 738 TSEFFELTSEELQIILTALCHRI---LMTYSVQDHMETRQOMSABLW--KERLAVLKEE 791

933 TQFOAQQPOQAAVLAFFVPHELNGSTLIINEIDKLTLESMSYKKNWIVEGRLLRLK-- 990
QY 792 NDKKAQKQKREMEAKENKENGKVENGLKTDKRRKRIKVFQOVTEADQMISAYKSRL 851
Db 991 --TVLAKTGRSEVEM-----GRPECLGRRSSRIMEETSQWEEBEEESAAVPGRR- 1042
QY 852 LAIQAKKEIEQEREMKV-KLEROQAEERIRKHAARAKAFOEGITAKAKLVNRRTPIGTD 910
Db 1043 ----GRDGEVDATASSPELERQIEKLSK-----QLFFKKLLHSSQMLRAVSLQD 1092
QY 911 RNHNRYWFLSDVPGLFIE-----KGWVHDSID----- 938
Db 1093 RYRRRYWVL-PYLAGIFVEGTGNLVPBEVIKKTDSLKVAHAASINPALFPMKMLAGS 1151
QY 939 -----YRFNHCKDHTVSGDEDYCPRSK----- 961
Db 1152 NTTASSPARARPLKTRPGFQPRHFKS-PVRGQDSEQQAQLQPEAQLHVPAQFPQL 1210
QY 962 -----KANLGN-----ASMTQHGTAATEVAVETTFKQG-- 991
Db 1211 QLQLOSHKGFLEBQESPLSLGQSDHLSQSAFLSWLSQTSQSHLSLSSVLTPDSQPKL 1270
QY 992 -----QNLWFLCDSQ----- 1001
Db 1271 DPAPQPPEEPDEAESPDLOAFWNISAQMPCNAAPTPLAVSEDOPTSPQOLASS 1330
QY 1002 ----- 1001
Db 1331 KPMNPSAANPCSPVQFSSTPLAGLAPKRRAGDPGMPQSTGLQPKRRGRPPSKFFKQ 1390
QY 1002 -----KELDELLNCLHPQIGRESQLEKREKYQDII 1033
Db 1391 MEQRYLTQTAQVPPEMCSGWWIWDPEMLDAMKALHPRGIRKALKHLNK-HRDFL 1449
QY 1034 HSIHL--ARKPNLGLKSCDNGOE-LLNFLRSLIEVATRLQKGGYGYEETSEFEARVIS 1090
Db 1450 QEVCLRPSADPIFEPRQLPAFQEGIMSNPK-----KTYETDLAVLQWVEELRQVIM 1503
QY 1091 -----LEKLKDPGECV----- 1101
Db 1504 SDOIRGWTCPSPDSTREDLAYCEHLSQSDITWRGPGREGIAPQRKTTNPLDLAVMRL 1563
QY 1102 IALOASVTKFQGMFAPKQRRKLQSDSAXTEEVDEEKKWVE-----BAK 1148
Db 1564 AALEONVRRYURELWP-----THEVLEKALISTPNGAPGTTTETISYE 1609
QY 1149 VASALEKWKTAIREAQTFSRMHVLLGMDACIKWDMSAENARCKVCPKKGDDKLILCDE 1208
Db 1610 ITPRIRIWRQTLQRCRSAHVCLCLGHLERSIAWEKSVNKTCLVCRKGDNDFFLLCDG 1669
QY 1209 CNKAFHLCLRALYEPDGEWQCPACQATARRNSGRNYTEESASEDSEDESEEE 1268
Db 1670 CDRGCHYCHRPMEAVPEGDFWCTVC-----LAQOVEGEFTQKRGF 1711
QY 1269 EEEEEEEDYEV--AGLRLPRKTRIGKHVIPPAAARSGRRPGCKKHSTERSOPKAPP 1325
Db 1712 PKRGQRKSGYSLNFSQDGRRRVLLKGRES--PAA--GPR-----YSEERLSPSK-- 1759
QY 1326 VDDAEVDELVLQTKRSRRQSLELOKCEBILHKYIRFSWPFREFTVRDEADYDVIT 1385
Db 1760 -----RRRLSMRNHSDLTFCETIILMEMESHDAAWFFXEPVNPRLVSGYRRIK 1808
QY 1386 HPMDFQTVQNKSCSGSYRSVQBFPLTDMKOVFTNAEYVNYCRGSHV 1429
Db 1809 NPMDFSTMRERLLRGYTSSEFAADALLVDFNCOTFNEBDDSEV 1852

RESULT 30
US-09-839-479-71
; Sequence 71, Application US/09839479
; Publication No. US20020039779A1
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.

; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042002
; CURRENT APPLICATION NUMBER: US/09/839,479
; CURRENT FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 71
; LENGTH: 1969
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-839-479-71

Query Match 6.4%; Score 509.5; DB 9; Length 1969;
Best Local Similarity 17.9%; Pred. No. 8e-20;
Matches 351; Conservative 247; Mismatches 604; Indels 763; Gaps 67;

QY 8 KPFLVNLPLGCEEPF-FTIPH-----TOEAFRTREYEYEARLERYSERIWTCKST 55
Db 189 KPLSLVNOAKETWYKLVPSDVLKAGNKNTSESSLTSELRSKREY-----KQA 241
QY 56 GSSQTHKEAWBEEQEVAAELK-----EEFPWVEKLVLEMVHHN-----TA 97
Db 242 PFSQLKQESSKSLKVIAALSNPKATSSSPA-HPKQTLNHNPNPFLTNALLGNHPNG 300
QY 98 SLEKLVDTAWLEIMTKYAVGECDFEV----- 124
Db 301 VIQSVIEAPLATTTKMQSKINENIAAASSTPFSSPVNLSTSGRRTPGNQTPVWPSAS 360
QY 125 -----GKEKML--KVIVKI-----HPL-----EKVDE 145
Db 361 PILHSQGEKAVNNVNPVKQHHSHPAKSLVEQFRGTDSDIPSSKDSNDEDEDE 420
QY 146 EATEKSDGACDPSKXENSSQIAQHQKETTIVVKEDEGRRESINDARRSPKLP7SL 205
Db 421 EDEDEDEDDSDSQSSEDSNSE--SDTEGSEEDDDKQDQSDSD----- 465
QY 206 KGERKWAAPPFLPHKYDVKLQNEDKIISNVPAOSLITERPPNKEIVRYFIRHNALRAG 265
Db 466 TEGE-----KTSMKLNKTTTS--SKSPMSLTCGHSFTPRNLHIK----- 501
QY 266 TGENAPWVEDELVKYSLPSKFSDFLLDPYKYMTLNPSTKRKNTGSDPRKPSKSK-TD 324
Db 502 ----APGSAPALCSESQSPA-----FLGTSSSTL---TSSPHSGTSKRRVTD 543
QY 325 NSSLSPLNPKLWCHVHLKSLSGSPLKVXKSNKSKSP-----EEHLEEMKXMSNP---- 376
Db 544 ERELRLPLEYG-WQRETRIRNFGG--RLQGEVAYYAPCGKKLRQYPEVYKILSRNGIMD 599
QY 377 -----KLHT-NFHIPKPGP-----PAKKPGK 396
Db 600 ISRDNFSFSAKIRVGDIFYEARDGPOEQMWCLLKEEDVIPRAMEGRGRPNPNRQAR 659
QY 397 HSDKPLKAKGRSKGI-----LNQKSTGNSKS 423
Db 660 EESRMRRRKGRPPNVGNAEFLDNADAKLLRKLQAEARQAQIKLLRKLQKQEQARVAKE 719
QY 424 PKK--GLKTPKTQMKQMTLLDMAKGTQKMTAPRNSGGTTPRTSSKPHKLPALHLYAY 481
Db 720 AKKQQAIIAAEBEKQKEQIKHKQEKIKRIQQ-----IRMEKEL--RAQOILEA 768
QY 482 YKENDKREDKBSALSCVISTKARLLSSEDRARLPEELSLVKQRYELLEHKKRWASMSSE 541
Db 769 KKKKKEE-----AANAKLLEAEKRIK-----EREMRRQAVLLKROER-----ER 808
QY 542 QRKEYLKKKBEELKKLKEKAKEREKEMLERLEKQKRYEQQEL---TGKNL----- 590

QY 377 -----KLHT-NFHIPKGP-----PAKPGK 396
Db 600 ISRDNFSFSAKIRVGDFEYARDGPQEQWCLLKEDVPIPRAMGRGRPPNPDRQAR 659
QY 397 HSDKPLKAKGRSGI-----LNGOKSTGNSK 423
Db 660 BBSRMRKGRPPNVGNABFLDNADAKLLRLKLOAQEAQAAQIKLLRLKLOQEQARVAKE 719
QY 424 PKK--GLKTPKTKMQMTLLDWAQTKQKTRAPRNSGGTPRTSSKPHKLPPAALHLIAY 481
Db 720 AKKQQAIAAAEERKQEQIKTHKQEKIKRIQQ-----IRMEKEL--RAQQIILEA 768
QY 482 KYENKOREDKRSALSCVTSKTKARLLSSDRAALPEELSLVOKRYELLEHEKKRWASMSSE 541
Db 769 KKKKEE-----AANKLLEAEKRIK--EREMRQQAVALKRQER-----ER 808
QY 542 QRKEYLKKREELKKLKEKAKERKEKMLEKQKRYEDEL--TGKNI----- 590
Db 809 RRQHMLMKAMEARKAEKERLKQEKDEKRLKERKLEQRLEMAKELKKPNEDMC 868
QY 591 -----PAFLVDTPB-GLPNTLFGDVAMVVEFLSCYGLLLPDAQYPTAVS-LMEALSA 643
Db 869 LADQKPLPELPRIGLVLVSGSTFSCLMVVOFLRNFVKVLGFDVNVDPNLSVLQEGILL 928
QY 644 DKGGFLYLNRVLVILLQTLQTLQDEIAEDY--GELGMLSEIPLTIHVSSELVRLCL 700
Db 929 NIGD--SMGEVDQLVRLLSAAVCDPGLITGYKAKTALGEHLNVGVNRDNNVSEILQIFM 986
QY 701 RRSVDQESSEGGTDDNDKSAAFEDNEVQDEFLEKLETSEFFELTSEBKLIQILTALCHRI 760
Db 987 -----EAHCGQTELTESLTKTKAFOAHTPAQK-AVLAFLINEL 1022
QY 761 LMTYSVDQHMETROQMSAELWELAVLKEENDKRAEKQKREKEMAKNKENGKVENGIG 820
Db 1023 ACKSVSVSEIDKNIDYMSNL-----RRDKNVVEGKLRIIHAKTGKRTSGG 1072
QY 821 -----KTDRKRIVKPEPOVDTEADMISAVKSRLLAQAOKKEREIQ 863
Db 1073 IDLGEQHPGLTPTGKRRRRKGGSDYDDDDDDSDO-----QGDEDDDE 1119
QY 864 E--REMKVQLERQABEERIRKHAAAEK-----AFQEGIAKAKLVMERTIG 908
Db 1120 EDKEQKGGKTDICEDEDEGQAASVEELEKQIEKLSQSQYRRKLPDASHSLASVMFG 1179
QY 909 TDRNHRNRYLWFDEVPGLFIE-----KGWH-----DSID 938
Db 1180 PDYRRRYWIL--PRCGGIFVGMESGEGLEETAKEREKLKKAESVQIKKEEMFETSGDSL 1238
QY 939 YRFNHC-----KDHT--VSGDEDYCPRSK-----KANLKNA-- 969
Db 1239 CSNTHCEQEKELKEKDNVNLFLQKPGSFSKLSKLEVAKMPPESEVMTPKNAGANGCT 1298
QY 970 -----SMNTOHGTAATEVAE-----TTTPKQG 991
Db 1299 LSYQNSGKHSLSGVQSTATQSNVEKADSNLFTGSSGKGFYPLPNDQLLKTILEKRN 1358
QY 992 QNLWFL-----CD----- 999
Db 1359 Q--WFSLLPRTCDTSLTHADMSTASLVTQSQPPSPSPPTAPLAGSSAQNPVGLNPF 1416
QY 1000 ----- 999
Db 1417 ALSPLOVKGVSMMGLQFCGWTGVVTSNIPPTLSVPSLGSGLGISEGNGSNFLTSNVA 1476
QY 1000 -----SQKELDELLNCLH 1012
Db 1477 SKSESPVQNERKATSAQPAAEVAKPVDPPSPKPIPEEMQFCGWRRIIDPEDLKALLKVLH 1536
QY 1013 POGIRESQKLEKRYQDIHSHIHLARKPNGLKSCDGNQBLNPLRSLDLIE--VATR 1069
Db 1537 LRGIREKALQKIQK-HLDYITQACLKKNK-DVAIIELNEES--NOVTRDIVENNSVEEQ 1592
QY 1070 LQKGGLYVEETSEFEARVISLEKLDQFCECVIALQASVIKKFLQGFMAPKQ----- 1121

Db 1593 AMEMDLSVLQVEDLERVAS-----ASLQ-----VKGWMCPEPASEREDLV 1634
QY 1122 -----KRRKQOSESAKTEEVDEKQWVE--EAKVA 1150
Db 1635 YFEHKSFTKLCKEHDEFTGEDSSAHLERKSDNPFLDIAVTRLADLERNIERRIEEDIA 1694
QY 1151 SALEKWKTAIBEAQOTFSRMHVLGLMDACIKWDSAEANARCKVCKPKKGGDDKLILCDECN 1210
Db 1695 PGLRVRRALSEARSAAQVALCIIQOLQSIKAWESIMKYCOICRKGNEEULLLDCDGD 1754
QY 1211 KAFHLFCLRLPALYEPDGEWQCPACQAPATA-----RRNSRGRNYTBESASE--D 1257
Db 1755 KGCHTYCHRPKITIPDGDWFCPACIAKASQGLTKLKLHVKGKKTNSKGGKVTLTGD 1814
QY 1258 SEDDESDEEBEEEEEDYEVAGLRPRKTIIRGKHVSIPPAARSGRRPGKPHSTR 1317
Db 1815 TEDEDS-----ASTSSSLKRGNKDLOQR 1837
QY 1318 RSQKAPVDDAEVDDELVLOTKRSSRSQSLQKCEEILHKIVKYRFSMPREPVPTRDEA 1377
Db 1838 KMEENTS--INLSKQESFTSVKPKRDDSKDOLALCSMLTETMETHEDAWPFLLPVNLKLV 1895
QY 1378 EDYDVVITHPMDFOTVQNKSCGSYSRVSQVEFLTDMKQVFTNAEYV 1422
Db 1896 PGYKVKIKPMDFFSTIREKLSGGQYPNLETALDVLVDFDNCETF 1940

RESULT 32
US-10-702-148-71
; Sequence 71, Application US/10702148
; Publication No. US20040063145A1
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042002
; CURRENT APPLICATION NUMBER: US/10/702,148
; CURRENT FILING DATE: 2003-11-05
; PRIOR APPLICATION NUMBER: PRIOR APPLICATION NUMBER: US/09/839,479
; PRIOR FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 71
; LENGTH: 1969
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-702-148-71

Query Match 6.4%; Score 509.5; DB 15; Length 1969;
Best Local Similarity 17.9%; Pred. No. 8e-20;
Matches 351; Conservative 247; Mismatches 604; Indels 763; Gaps 67;

QY 8 KPFLVNLPGEEPP-FTIPH-----TQEAFTREEYEARELRYSERIWTCKST 55
Db 189 KPLSLVNOAKKETYMKLVPSPDVLKAGNKVTSESSLTSELRSKREQY-----KQA 241
QY 56 GSSQLTHKEAWEEEOEVAELLK-----EFPFAWTEKLVLEMVHHN-----TA 97
Db 242 FPSQLKKQESSKSLKVIAALSNPKATSSSPA-HFKQTLNHNHPFLTNALLGNHPNG 300
QY 98 SLEKLVDTAWLEIMTKYAVGEECDPEV----- 124
Db 301 VIQSVIQEAPLALTTTKQKSKINENIAAASSTPSPVNLSTSGRTPGNTGQTPVMPAS 360
QY 125 -----GKEKML--KVKIVKI-----HPL-----EKVDE 145

Db 361 PILHSQKEKAVANNVNVKTOHHSPAKSLVBFQRTGDTSDIPSSKSDSDSEDEEEDDE 420
Qy 146 EATEKESDGCADSPSDKENSIOADHOKKETTWKVEDKERRSINDRARRSPKLPSTL 205
Db 421 EDEDEDEDDSDQSDSDNSE--SDTEGSEEDDDKQDSDSD----- 465
Qy 206 KGERKWAAPKFLPHKYVVKVQWEDKIIISNVPADSLIRTPPNKSIYFIRHNALRAG 265
Db 466 TEGE-----KTSMLNKTTSS--SKSPMSLTGHTSPRNLHIK----- 501
Qy 266 TGENAPVWVEDELAVKYSLSKPSDFLLDPKYMTLNPSTKRNKNTGSPDRKPSKSK-TD 324
Db 502 ----AFGSPAALCSQSPA-----FLGTSSSTL---TSPHSGTSKRRVTD 543
Qy 325 NSSLSPLNPKWCHVHLKKSLSGSLPKVKNKSNKSP---EEHLEEMMMKSPN----- 376
Db 544 ERELRLPLEVG--WQRETRIRNFG--RLQGEVAYVAPCGKLRQYPEVIKILSRNGIMD 599
Qy 377 -----KLHT-NPHIPKKGK-----PAKPKGK 396
Db 600 ISRDNFSAKIRVGDYFYEARDGPQBMQWCLLKEDVIPRIRAMEGRGRPPNPDRQR 659
Qy 397 HSDKPLKAKGRSKI-----LNQKSTGNSKS 423
Db 660 EESRMRRKGRPPNVGNVNAEFLDNADAKLRKLOAQBARQAQIKLLRKLQKQBARVAKE 719
Qy 424 PKK--GLTKPTKWKQMTLLDMAKGTQKMTAPRNSGGTPTSSKPKHLPPAALHLIAV 481
Db 720 AKQQAIAAAEKQKQEKIKHQEKIKRIQ-----IRNEKEL--RAQQILAE 768
Qy 482 YKENDREDKRSALSCVISTARLLSSEDRARPEELRSLVOKRYELLHKKRWASMSBE 541
Db 769 KKKKEE-----AANAKULEAEKRIK-----EREMRQQAVALLKQER-----ER 808
Qy 542 QRKEYLKKKEELKKLKEKAKERREKEMLERLEKOKRYEDQEL-----TGKNL----- 590
Db 809 RQHMMLKAMEARKAEKELKQEKREKLNKERKLEQRLELEMAKELKKPNEDMC 868
Qy 591 ----PAFLVDTPPE--GLPNTLPDGVAMVVEFLSCYSGLLPDAQYPIITAVS--LMEALSA 643
Db 869 LADQKPLPELPRIPLGLVSGTSDCLMVVQVLFNFKGVLFVDVNLVQLQEGILL 928
Qy 644 DKGGFYLNARVILLQTLLOLLODEIAEDY---GELGMKLSIEPLTLHSVSELVRLCL 700
Db 929 NIGD--SMGEVOLLVRLLSAAVCDPLITGYKAKTALGHEHLLNVGNRDNVSEILQIFM 986
Qy 701 RRSVDQSESGSDTDNKSAAEFEDNEVQDEFLEKLETSEFFELTSEEKLOLITLALCHRI 760
Db 987 -----EAHCGQTELTSIKTKAFAQHTPAQK-AVLAFLINEL 1022
Qy 761 LMTYSVDHMETRQMSAELWKERLAVLKEENDKKRAEKQKREMEAKKNKENGKVENGLG 820
Db 1023 ACSKSVSEIDKNIDYMSNL-----RRDNVVEGLKRLRIIHAKKTKGRTSGG 1072
Qy 821 -----KTRDKRRIVKFPFQVDTAEADMISAVKSRRLLAIAQAKKERIQ 863
Db 1073 IDLGEQHPGLGTPPKRKRKRGDSDYDDDDDDSD-----QDEDEDEDE 1119
Qy 864 E--REMKVKLERQAEERIRKHAABK-----AFQEGIAKLVMRTPTG 908
Db 1120 EDKEDQKGGKTDCEDEDEGDQAAVSELEKQTEKLSKQSQVRRKLPDASHLSRSMVG 1179
Qy 909 TDRNHRNYLFSDEVPLFTE-----KGWVH-----DSTD 938
Db 1180 PRYRRRYWIL-PRCGGIFVEGMESEGLEEEIAKEREKUKAASVOIKSEMETSQDLSN 1238
Qy 939 YRFNHC-----KDHT-----VSGDEDYCPRSK-----KANLGNA-- 969
Db 1239 CSNTDACEQKEDLKEKDNINLFLQKFGSKLSKLLLEKVAKMPSEWTFPKNAGANGCT 1298
Qy 970 -----SMNTQHTATEVAE-----TTTTPKQK 991

Db 1299 LSYQNSKGHSLGSVQSTQTSNVEKADSNLNTGSSGPKGFYSPLPNDQLLTKLTKNKR 1358
Qy 992 QNLWFL-----CD----- 999
Db 1359 Q--WFSLLPRTCDTSLTHADMSTASLVTPOSQPPSKSPPTPAPLGSSAQNPVGLNPF 1416
Qy 1000 ----- 999
Db 1417 ALSPLQVKGGSVMGLOFCGWPTGVVTSNIPFTLSVPSLSGLSGLSEGNSSPLTSNVAS 1476
Qy 1000 -----SQKELDELINCLH 1012
Db 1477 SKSESPVPONEKATSAQPAAEVAKPVPDPPSPKPIPEEMQFGWMRIIDDEPKALLKVLH 1536
Qy 1013 PQGIERSQLKERLEKRYQDIIHSIHLARPNLGLKSCDGNQELLNFRSLDIE---VATR 1069
Db 1537 LGIREKALQKQIK-HLDYITQACLKNK-DVAIIELNENEE--NQVTRDIVENWSVEEQ 1592
Qy 1070 LQKGGLYVEETSBEFEARVISLEKLDKDFECVIALQASVIKKPQGFMAPKQ----- 1121
Db 1593 AMEMDLVLQVVEDLERRVAS-----ASLQ-----VKGWMCPEPASEREDLV 1634
Qy 1122 -----KRRKLOSEDSAKTEEVDEEKKWVE--BAKVA 1150
Db 1635 YFEHKSFTKLXEHGDEFTGEDSSAHLERSNDPLDIAVTLADLERNIERRIEEDIA 1694
Qy 1151 SALEKWKTAIREAQTFSRMHVLLGMLDACIKWMSAENARCKVCPKKGDDDKLILCDECN 1210
Db 1695 PGLRVRRALSEARSAAQVALCIIQLOKSIANEKSIKMYVCQICRKGDNBEULLLDCGCD 1754
Qy 1211 KAFHFLCRLPALYVEPDGWEQCPACOPATA-----RNSRGRNVTESASE-----D 1257
Db 1755 KGCHTYCHRPKITTPDGDWDFCPACITAKASGGTLLIKLHLVKGKKTNESKGGKVTLTGD 1814
Qy 1258 SEDDSDEEBEEEEEEEDYEVAGLRRLRPKRTIRGKHSVIPPAAARSGRPGKKPHSTR 1317
Db 1815 TEDEDS-----ASTSSLKRGKNKOLQKR 1937
Qy 1318 RSQKRAPPVDDAEVDELVLQTKRSRRRQSLQKCEEILHKIVKYRFSWPPREPVTTRDEA 1377
Db 1838 KNEENTS--INLSKQESFTSVKPKKEDDSKDLALCSMLITEMETHEDANPFLPVNKL 1895
Qy 1378 EDYDVITHMPFQTVQNKSCGVSYSRVOEFLTDMKQVFTNAEYV 1422
Db 1896 PGYKKVKKPMPDFSTIREKLSGGQYNLETFTALDVLRLVFDNCETF 1940

RESULT 33
US-09-864-761-34543
; Sequence 34543, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aemica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30

```
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 34543
; LENGTH: 65
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC005089.2
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.73
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 0.78
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 2.2
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2.3
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.3
; OTHER INFORMATION: SWISSPROT HIT: Q84133, EVALUATE 5.20e-01
; OTHER INFORMATION: EST_HUMAN HIT: BE258324.1, EVALUATE 3.00e-32
US-09-864-761-34543

Query Match 4.3%; Score 341; DB 9; Length 65;
Best Local Similarity 100.0%; Pred. No. 3.7e-12;
Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 233 ISNVPADSLIRTPPNKEIVYFIRHNALRAGTGENAPWVVEDELVKYSLPSKFSDFL 292
Db 1 ISNVPADSLIRTPPNKEIVYFIRHNALRAGTGENAPWVVEDELVKYSLPSKFSDFL 60

QY 293 LDPYK 297
Db 61 LDPYK 65

RESULT 34
US-09-864-761-34544
; Sequence 34544, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
```

```
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 34544
; LENGTH: 69
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC005089.2
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.9
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.8
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2.5
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.4
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.72
; OTHER INFORMATION: EST_HUMAN HIT: A1498634.1, EVALUATE 5.00e-32
; OTHER INFORMATION: SWISSPROT HIT: P13816, EVALUATE 1.70e-01
US-09-864-761-34544

Query Match 4.2%; Score 338; DB 9; Length 69;
Best Local Similarity 98.5%; Pred. No. 5.8e-12;
Matches 67; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 124 VGKEMKLVKIVKIHPLKVDDEATEKKSDGACDPSDDKENSQIAODHQKKTWVKED 183
Db 1 VGKEMKLVKIVKIHPLKVDDEATEKKSDGACDPSDDKENSQIAODHQKKTWVKED 60

QY 184 EGRRESIN 191
Db 61 EGRRESIS 68

RESULT 35
US-10-369-493-1586
```

; Sequence 1586, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 1586
; LENGTH: 1790
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-10-369-493-1586

Query Match 4.0%; Score 322; DB 15; Length 1790;
Best Local Similarity 20.2%; Pred. No. 2.6e-09;
Matches 274; Conservative 222; Mismatches 471; Indels 392; Gaps 58;

QY 17 PGEPEFTPIPTQEAFTREYERLERYSERIWTCKSTGSSQUTHKEAWE-----EE 69
DB 719 PDEPINKI-----SPEEVEKLORQCTKLGEITSLOT--ETESTHENLTKELIATNEH 771
QY 70 QEVAE-----LLKEFFPAWYKLVLEVMVHNHTA-SLEKLV-----DT 105
DB 772 KELDEKYQILNSHSHSILKENFS-----ILETELKNVRDSDLDMTQLRDVLETKKENOT 825
QY 106 AWEIMTKVAVGEC--DFEVGKMKLVKIVKIHLEKV-----DEATEKKSOG 154
DB 826 ALLEYSTHIHQDSITKLEKGLLETILSQKKKAEDGINKMGKDLFALSREMQAVERN--- 882
QY 155 ACDSPSDKNSQIAQHQKBTVVKEDEGRESINDRARRSPKLPPTSLLKGERKWAP 214
DB 883 -CKNLQEKDKSN--VNHQETYSLKEDIAAIT-----913
QY 215 PKPLPHKYDKLQNEDKIISNPADSLIRTERPNKEIVRYFIRHNALRAGTGENAPWV 274
DB 914 -----BIKAINENLEEMKIQCNNLSKEKEHISKELVEYKSRFQS-----952
QY 275 EDELVKYSLPSKFSDFLLDPYKWTILNPSTKTKNTGSPDRKPSKSKTDNSLSPLNP 334
DB 953 HDNLVAK--LTEKLKS-LANNYKDMQAEENSLIKAV-----EESKNESSTQLSLNQLN 1001
QY 335 KLWCHVHLKKSLSGSPKLVKNSKNSK-----SPEEHLEEMKMMSPNKLHTNFHPPKGGP 390
DB 1002 KI-----DSMB-----QEKENFOIERSIEKNIEQLKKTISDLE-QTKBEIISKDS 1047
QY 391 AKKPGHSDKPLKAKRSGKILNGQKSTGNSKSPKGLKTPKTKMQLTLLDMAKGTQM 450
DB 1048 SKDEYESQISLLKEKET-----ATTANDENVNKISELTKTREBLEAEALAYKNL--- 1097
QY 451 TRAPRNSGGTP-RTSSKPHKHLPPAALHLIAYKENKDRDKRSALSCVISTARTLSS- 508
DB 1098 ----KNELETKLTSEKALKEVENEHL-----KEEKIQLEKEA-----TETKOQNSL 1143
QY 509 ----EDRALPELRSLAVOKRYELLHKKRWAS-----MSBEORKEYLKKREEL 554
DB 1144 RANLESLEKEHEDLAQLKYEIQIANKERQYNEEISQUNDEITSTQQENESIKKKNDEL 1203
QY 555 KKKLKEKAKERREKEMLEKRYEDQELTGKPLPAFLVDTPEGLPNTLFGDVAMVV 614
DB 1204 -----EGEVKAMKSTSEQSNLKKSEIDALNLQIKELKKKNETNEASLESIKS-V 1253
QY 615 EFTVCSYGLLLPDAQYPIITAVSLME---ALSADKGGFLYNRVILVILLOTLQTLLODEI 671
DB 1254 ESETVVIKELQDCNFCNEKVEVSELEDKJLASEDKNS-KYLE-----LQKES 1298

QY 672 AEDYGELGMKLSSEIPLTLHSVSELVRLCLRRSDVOBESESGDTDDKDSAAFEDNEVQDE 731
DB 1299 EKIKEELDATTTELKIQLEKITNL-----SKAKESESELRLKKTSS--SERKNAEE 1349
QY 732 FLEKLETS-EFFELTSEEKLOILTALCHRLIMTYSVQDHMETRQOMSAELKWERLAVLKE 790
DB 1350 QLEKLKNEIQIKNOAFERKRLNNEGSSITTOEYS--EKINTLE-----DELRLQN 1399
QY 791 ENDKKRAE-KQKREKMEAKNKGKVENGLGKTRKRVKPEQVQDTEADEMISAVKSR 849
DB 1400 ENELKAKEIDNTRSELEKVSLSNDEL-----LEEKONTIK-----SLQDEILSYDK 1446
QY 850 -----RLLAIAQAKKEREIQE-----REMKVKLERQAE--EERIRKHAAAEEKAFQE 893
DB 1447 ITRNDEKLJLSIERDNKRDLESKEQLRAAQESKAKVEEGLKLEESSSEKAELE----- 1501
QY 894 GIAKAKLVMRPTPIGTDNRHNRVWLFSDVPLGFIKGGVWHDSDYRPNHHCXDHVTSGD 953
DB 1502 ---KSKEMMKLESTIESNETE-----LKSSMB-----TIRKS 1531
QY 954 EDYCPRESKANLGKNASMTQHGTADEVAVETTPPKQGNLWFLCDSQKELDELLNCLHP 1013
DB 1532 DEKLEQSKKS--AEEDIKNLQH-----EKSLIRINSEKIDELKSKLRI 1576
QY 1014 QGIRESQLERLEKRYQDIIHSIHLARKPNLGLKSCDGNQELNFLRSLDILIEVATRLQKG 1073
DB 1577 EAKSGSEL-ETVQELANNAQEKIRINABENTVLKS-----KLEDIE----- 1616
QY 1074 GLGYVEETSEFPAVLSLEKLDGFCVIALQASVIAKFKFLOGFMAPQKRRKLOSEDSAK 1133
DB 1617 -----RELKDQAEIKSNQEEKE-----LLTSRLKELEQELDSTQQAQKSEERRAE 1664
QY 1134 TREVDEEKKVBEAKVASALE-KWKTAIRAETQFSR-----MHVLLGMLD 1177
DB 1665 VRKFQVEKSQLDEK--AMLESTKYNDLVNKRQAWKQDEBTVKTTTSQSQOEIEKJLAKELD 1722
QY 1178 ACIKWDMSAENARCKVC--PKKGEDDKLILCDECNKAPHFLCLRLPALYEPDGEWQCPAC 1235
DB 1723 ----NLKAENSKLKEANEDRSIEDDLMLLVTDLDE-----1753
QY 1236 QPATARRNRGRNYTEESASESDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDE 1274
DB 1754 ----KNAKYRSKLKDLGVEISSDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDE 1786

RESULT 36
US-09-839-479-67
; Sequence 67, Application US/09839479
; Publication No. US20020039779A1
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042002
; CURRENT APPLICATION NUMBER: US/09/839,479
; CURRENT FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 67
; LENGTH: 65
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-839-479-67

```
Best Local Similarity 98.5%; Pred. No. 1.8e-10; Mismatches 0; Indels 0; Gaps 0;
Matches 64; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 630 YPITAVSLMEALSADKGGFLYNRVLLVILLOTLQDEIAEDYAGELGKLSIPLTL 689
Db 1 YPITAVSLMEALSADKGGFLYNRVLLVILLOTLQDEIAEDYAGELGKLSIPLTL 60

QY 690 HSVSE 694
Db 61 HSVSE 65

RESULT 37
US-10-376-537-68
; Sequence 68, Application US/10376537
; Publication No. US20030224405A1
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042001
; CURRENT APPLICATION NUMBER: US/10/376,537
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US/09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 68
; LENGTH: 65
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-376-537-68

Query Match 3.9%; Score 311; DB 15; Length 65;
Best Local Similarity 98.5%; Pred. No. 1.8e-10;
Matches 64; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 630 YPITAVSLMEALSADKGGFLYNRVLLVILLOTLQDEIAEDYAGELGKLSIPLTL 689
Db 1 YPITAVSLMEALSADKGGFLYNRVLLVILLOTLQDEIAEDYAGELGKLSIPLTL 60

QY 690 HSVSE 694
Db 61 HSVSE 65

RESULT 38
US-10-702-148-67
; Sequence 67, Application US/10702148
; Publication No. US20040063145A1
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042002
; CURRENT APPLICATION NUMBER: US/10/702,148
; CURRENT FILING DATE: 2003-11-05
; PRIOR APPLICATION NUMBER: PRIOR APPLICATION NUMBER: US/09/839,479
; PRIOR FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
```

```
; SEQ ID NO 67
; LENGTH: 65
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-702-148-67

Query Match 3.9%; Score 311; DB 15; Length 65;
Best Local Similarity 98.5%; Pred. No. 1.8e-10; Mismatches 0; Indels 0; Gaps 0;
Matches 64; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 630 YPITAVSLMEALSADKGGFLYNRVLLVILLOTLQDEIAEDYAGELGKLSIPLTL 689
Db 1 YPITAVSLMEALSADKGGFLYNRVLLVILLOTLQDEIAEDYAGELGKLSIPLTL 60

QY 690 HSVSE 694
Db 61 HSVSE 65

RESULT 39
US-10-408-765A-1215
; Sequence 1215, Application US/10408765A
; Publication No. US20040101874A1
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fahy, Eoin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; APPLICANT: Warnock, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; TITLE OF INVENTION: IDENTIFIED IN THE MITOCHONDRIAL PROTEOME
; FILE REFERENCE: 660088.465
; CURRENT APPLICATION NUMBER: US/10/408,765A
; CURRENT FILING DATE: 2003-04-04
; NUMBER OF SEQ ID NOS: 3077
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1215
; LENGTH: 1327
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-408-765A-1215

Query Match 3.8%; Score 305.5; DB 16; Length 1327;
Best Local Similarity 21.0%; Pred. No. 1.5e-08;
Matches 323; Conservative 208; Mismatches 517; Indels 491; Gaps 76;

QY 28 TOEAPRTREEYEARLER-----YSERIWTCKSGSSQLTHKEAWEEBEEQVAELLKEEF 80
Db 48 TLEQEEKEQFFARLEKGLTSSIDYSR---LNKELDSNDSTHFKALHNSQANAEITDDE- 103

QY 81 PAWYEKLVMVHNHTASLEKLVDVTAWLEIMTKYAVGEECDPE---VGKMKMLKVIVKI 137
Db 104 -----HENESKHEEL-----AENYSDDDFEDYVGA----- 128

QY 138 HPLEKVDDEATEKKS DGA--CDSPSSDKENSQIAQ-----DHQKKTETVVKDEGRRESIN 191
Db 129 -PLTTKDEMPKSKENSKSEKISVPKQBEKTMGLANVLLDLSLDSVAEVLNDE-----Q 181

QY 192 DRARRSPKLPSTLKKGERKWAPPKFLPHKYDVK-----LONED--KIISNVP 237
Db 182 DKITPKPRCLP-EMTENEMTGTGVSQSSDDEALHQAHCIAHSLGDEDKQKTESNTV 240

QY 238 AD--SLITERPPNKEIVRYFIRHNALRAGTCENAPWVVEDELVKYSLPS-KFSDFLLD 294
Db 241 EDIKSVKGHQENEE-----NSKNISTWESDLPTVE-ELMKPIRIDSFGISGFDLQ 291

QY 295 PKYVWTLNPNPSTKRNKWTGSPDRKPKSKSKTDSNSSLSPINPKLWC-----HVHLKLSLSSGP 350
Db 292 PV-----SSEKVAERKETETFFSSLPKKNPNILSDSQHVNLFDDKNDEN 336

QY 351 LKVKNSKNSK-----SPEEHLEEM-----MKWMSPNKLHTNFH-- 383
```


337 VILQKTTNEMSCFQVTEVTATEEHVDKMYLNLKTKITVNSSLSQDDKINKTYRSQ 396
384 -----IPKGPAPKPKGKSDKPLKAKGRSKGLNGOKSTGNSKSPKKGLTKPTKTM 435
397 LSSEEGAVNGKQVPYK--ARAPLLKPKQSGLYASVRSSGYK--PSSPLMFST-L 452
436 KQMTLDMAGTQKMTAPRNSGGTPTRTSKPHK-----LPPAALHLIAYYKEN-- 485
453 EKKTSBIIKSKNLRIS-----TSNQPKKBEILSGTKLIPKALDKPAHKTESCL 503
486 ---KOREDKSALSVCWIKTARI--LSSEDPARLPBELRSLVQRYELHELHKGKWSMSE 541
504 STRKXSNPTETDSCIQFQDTSIGYGENK-----EKKLIMFKRVQEAEDKWRGAQALIE 558
542 QRKEYLKKREELKKLKEKAKERRKEMLERLEKQVEDQELTKNLPAPFLVDTPEG 601
559 QIKATSEKELENK--BELKKQKEKELF-----KLNQDNVILQAKLSSFE--ETNKK 609
602 LPNTLFGDVMVVEFLSCYSGLLPDAQYPTITAVSLMEALSADKGGFLYLNRLVILQT 661
610 QRLWLFGEAA-----DPVTGKELKQ-----IQK 632
662 LIQTLLODEIAEDYGBELGKWLSEIPLTHSVSELVRLCLRRSDVQSESGSDTDDNKSA 721
633 EIQE--QETLLOGYQ-----ENERLYNQVKDLOEQ----- 661
722 APEDNEVQDEFLEKLETSBEFFELTSEKLIQTLALCHRI--LMTYSVODHMETRQMSAEL 780
662 ---NKNBERMPKENSJFSEVAS-----LKEQMHKGRFLSQVVEDSEPTNRNQFTDL 711
781 WKERLAVLKEEN---DKRAEKQKQK--EMEAKNKENGKVENGLG----- 820
712 LAELRMAQEKESLLEDIKRLQDKQALEVDPEKMKERDQAKQIAYVTGKEYLKIIL 771
821 KTRDKRIVKFPQVDEADMSIAVKSRRLLAIAQAKREIREIQEREMKVLE---RQAE- 876
772 EETHKOEISRLQKRLQWYAE-----NQELDKDALRLREANEETEKLKLEITEKLAES 824
877 ---EERIR--KHKAARAEKAFQ-----EGIAKAKLVMBRTPIGTDNRNRYWLFSD 921
825 GNPSIRQKIRLKDAAKAKIQDLERQVKEMEGILK-----RYP-----N 865
922 EYVPLFIEKGWHDSDIDYRNFHHCKDHTVSGDSDYCPRSKK--ANI--GKN-----ASMNTQ 974
866 SLPALILAAAGDITVD-----KNTVEFME---KRILKLEADLEGKDEDAKKSLRTM 914
975 HGTATEVAVETTPPKQGNLWFLC-----DS-----QKLEDELNCLHPQGIRESQLKE 1023
915 EQQFQWKIQYEQRLRQEQQLACKLNQHDSPRIKALEKELODIIKEA-HQITVR----- 967
1024 RLEKRYQDIHSHLARKNGLKSCDGNQELLNFLRSLDIEVATRLQKGLGYVEETSE 1083
968 NLEAEIDVLKH-----QNAELDVKKNDKODEDFQ-----SIEFQVEQAH 1006
1084 FEARVISL-----EKLKDFGEVIALQASVIKKFLOQFMAPKQRRKLOESDSAKTE 1135
1007 AKAKVLRLNEELAAKBEIQDLSKTVRLQ-----KDRMMLSNQSKGRE 1052
1136 EVDEE--KQWBEAKVASALE---KWKTAIREAQTFSRMHVLLGMLDACIKWDMSAENAR 1190
1053 EMSAKRAKKDLVHSSKGNANSFPGTLDLSKLYQHTFTDGHVS-----EVLQENYR 1102
1191 CK-----VCPKGEDDKLILDEC-----NKAHFLCLRLPALVEPDPGEWO 1231
1103 LKNELEGLISEKNE---LKKKGEAVNQPNFNSMRRVKEDTAAHIASLKAS-----HQRETE 1155
1232 CPACQAPATARRNRRGNRYTEEGASED-----SEDESEDEEEE---EEEEEEEDYEV 1281
1156 KLLCQNAVENSCKVAELNRKATQEVLRHFPQSQVNELOSKQESLVVSEVEEILQKEI 1215
1282 AGL--RLRPRK---TIRGKHSHVPPAARGRRPGKPHSTRSQPKAPPVDDAEVDLVL 1336

Db 1216 TKLLEELREAKENHTPEMKHFV-----GLEKKIKQMEMRHAQ-----REQELOQIIQ 1262
QY 1337 QT-----KRSSRRQSLEQKCEIILHKIV 1360
Db 1263 QTHQVVEQONKEVEKWKRLAQLKQRELEKFRTELDL 1301
RESULT 40
US-10-408-765A-254
; Sequence 254, Application US/10408765A
; Publication No. US20040101874A1
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fahy, Eoin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; APPLICANT: Warnock, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; IDENTIFIED IN THE MITOCHONDRIAL PROTEOME
; FILE REFERENCE: 660088.465
; CURRENT APPLICATION NUMBER: US/10/408,765A
; CURRENT FILING DATE: 2003-04-04
; NUMBER OF SEQ ID NOS: 3077
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 254
; LENGTH: 3225
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-408-765A-254

Query Match 3.8%; Score 302.5; DB 16; Length 3225;
Best Local Similarity 18.5%; Pred.No.6.8e-68; Indels 661; Gaps 80;
Matches 358; Conservative 279; Mismatches 633;
QY 30 EAFRTREEEARLERYSERIWTCKSTG-----SSQUTHKEAVEEEOEVALKEEFPFA-- 82
Db 997 ETERGEVEDKENKEYSEKCVTSKQOEIYILKQITISEKV--ELQHIRKDLKEELAAEE 1054
QY 83 WYELKLVLEM-----VHNTASLEKLV-----DTAWLEIMTKYAV 116
Db 1055 QFQALVKQMNQTLQDKTNQIDLLQABISENQAIQKLTISNTDASDGSVAL--VKETVVI 1113
QY 117 GEECDFEVG-----KEKMLKVKIKVHPLEKVDDEATEKKS--DGACDSPSDKEN 165
Db 1114 SPC---TGSSEHWPELEEKILALEKEKEQOKLQLEALTSRKAILKKAQEKERHREE 1170
QY 166 SSQIAQDHQKQKTVVKEDEGRRESINDRARRSPRLPTSLKKGKRWAPPKFLPHKYDVK 225
Db 1171 LKQKDDYNRLQEQFDEQSKENENIGDQLRQLQIQVRESI-----DGK 1213
QY 226 LQNEED--KTIISNVA-----DSLIRTPPNKEIVRYFIRHNALRAG 265
Db 1214 LPSTDQOESCSSTPGLEEPFRKATQHHTQPVLESNLCPDWPSHSEDA-----SALQGG 1267
QY 266 TG-ENAPWVEDELVKYSLPSKFSDFLLDPKYMTLPSTKTKNTGSPDRKPSKSKTD 324
Db 1268 TSVAQIKAKLKEIEAKVELELVSS-----TTSELTKKSEEVFQIQOINQKQGLE 1318
QY 325 NSSLSPLNPLKCHVH---LKKSLSGSPKLVKNSKNSKSPBEHLEEMKMMKSPNK----- 377
Db 1319 ISLKTVSHE---AEVHAESLQKLESSQLQIAGLEHLRELQKULDELQKLSKKEEDVS 1375
QY 378 -----LHNTFHIHPKGPAPKPKGKHSK----- 400
Db 1376 YLSGQLSKEAALTKIQTEIIEQEDLIKALHTQLEMQAK-----EHERIKQLQVEL 1427
QY 401 -PLKAKGRSKGLNGOKSTGNSK-----SPKKGKTKTKTKMKTLLDMAGT-QKMTFR 452
Db 1428 CBMKQPEIGESRAKQIQKQQAALISRKALKENKSKLOEELS-----ARGTIERLTK 1484
QY 453 APRNSGGTPTRTSSKPHKLPPAALHLIAYYKENKOR-----EDKRSALSCVISKT 502

Db 222 RSTANVLBEETTIVKKEDEKELVK-----LPVIVKLE-----KPLPE----- 258
Qy 291 FLLDPYKWTNLNPSYKRNKNTGSPDRKPSKSKSTDNSSLSPPNPKLWCHVHLKXLSGSP 350
Db 259 ---NEEKIIKEESDFKENVKPIKVEKCEKADPKDTS-----SMEKDVQEP 305
Qy 351 LKKVKNKSKS---PEHLEEMKMGSPN-----KLHNEHPPKGP----- 389
Db 306 ERIFEGGNKSHSHEITEKSLKNDQQAQKIPLKREIKLSDDFSPVKGPCKSVT 365
Qy 390 PAKKPGKSHDK-----PLKAG-RSKGILNGKS-----TGNKSPKKGLTKPTK 434
Db 366 PTKFELKDEIKQEBETCKRISITITALGHEGQLVNGEVSDEVAFNPKEPIETKFYETK 425
Qy 435 -----MQMTLLDMAKGTQKTRAPRNSGGTPRTSSKPHKHLPPAALHLIAYKENKDR 488
Db 426 EESYSPSKDRNIITEGNGTESI-----NSVITSMKTGELKETAPL-----R 467
Qy 489 EDKRALSAC--VISTARLLSSDRARLPEELRSIVOKRYELLEHKRWASMESEQRKEY 546
Db 468 KDADSSIVLEIHSQAKIIEPD-----PPMETSLDSS-----EMAKLSSKLTALSSTES 518
Qy 547 LKKKEELKXKLKEXAKERRER--EMLERLEKQKRYEDQELTGKNLPAFRLVDTPEGLP- 603
Db 519 CTMKGEESPKTK---KOKRPPILSCLEKESKK-----TFLDKDAQRLSPPIPEVPK 569
Qy 604 NTLFGDVAMVEFLSCYGLLLPDQAQYPIITAVSLMEALSADKGGFLYLNRLVILLQTL 663
Db 570 STLESEKPGSP-----AAETSPSPNIIDHCEKLAKE----- 602
Qy 664 OTLLQDEIADYGEIGMKLSIPL--TLHSVELVLCRLRRSDVQES---EGSDTDNDKD 719
Db 603 -EVSQCSTSTVG--GQSVKVKVDLTLKEDSEFTKVMNDLNAQTGIESPSETKGMQ 659
Qy 720 SAAFDNEVDQFLEKLETSEFFELTSBEKLIQILTALCHRLMTYSVQDHMETROQMSAE 779
Db 660 KSKFYKLVPE---BETTASENTEITSRQXE-----GIKLTIRISRRKKKPSPP-- 707
Qy 780 LMKERLAVLKEBNDKRAEKQKRWEMAKNKGKVENGLKTDKRRKRVKPEPOVDTEA 839
Db 708 -----KVLPEPKOETKEBEK-----TNVGRTLAR-----SPRI-SRP 741
Qy 840 EDMISAVKSRLLAIQAKERIQEREMKVKLERQAEERIRKHKAAAEKAFQEGIAKAK 899
Db 742 TAKVAEIRDQK--ADKKGEGEVEEESTALQKTDKKEILKK-----SEKOTNSKVKVK 795
Qy 900 LVMRRTPIGT-----DRHNRYLFP--DEVFGLFIEKGWVHDSIDYRFRHCKDHTVSG 952
Db 796 -----PKGKVRWTGSRTRGR--WKYSNDESEGGSGSEKS--SAASEEESEKSEAILAD 846
Qy 953 BDDYCPRSKKANLGNKNA-----SMNTQHGATGATEVAVETTPKQGNLWF-----L 997
Db 847 DDEPC---KKCGLPNHPHELILLCDSGSHYTAACLRPLMIIP-DGE--WFCPPCQHKLL 900
Qy 998 CDS-QKELDELNLCHLPHQIGRESQKLERKRYQDIH-----SIHLARKPNLG---L 1046
Db 901 CEKLEQLODLVAL-----KKERAERKRLVYVGISIEINIIPQBPDPFSEDQEE 952
Qy 1047 KSCDGNQELLNPLRDLIEVATRLQK-----GGLYGVEE 1080
Db 953 KKKDSKSKKANLLE---RRSTRTRKCI-SYRDEFDEIDEAIEDDIKEADGGVGRGKD 1008
Qy 1081 TSEFARVISLEKLDQFGEVICALQASVIKPFLOGFMAPKQKRLKQESDSAKTEEVDEE 1140
Db 1009 IS-----TITGHRGDI--STILDEERKENKRPQRAAAARRKRRRLNLDLS--DSNLDEE 1060
Qy 1141 KQWVEAKVASALEKWKTAIREAQTFSRMHVLLGLMDACIKWMSAENARKVCPCKGED 1200
Db 1061 ES-EDFKISD-----GSQDEFV---VSDEN-----PDESEE 1089
Qy 1201 DKYLILCDE---CNKAFHFLCFLPALYEVDPGEMQCPACQAPATARNRSGRNYTESASE 1256
Db 1089 DPPSNDSDTDFCRRLRHPSRP-----MQSRRLRRTKPKKYS 1129

Qy 1257 DSEDDSEDEEEEEEEDYEVAGLRLPRKTIIRCKHSVI-----PPAARSGRP 1309
Db 1130 DDEEESEENRSESDSDFDDFDDFVETRRRRRRNQKQIINYKEDSESDGSKSLRR 1189
Qy 1310 GK---KPHSTRRSQKAPP---VDDAEVDELVLQTKRSSRRQSLELQKCEBI----- 1355
Db 1190 GKEIRRVKRLSSSESEESYLSKNSEDEDELAKESKRSVRKGRSTDEYSEADEEESEE 1249
Qy 1356 -----LHKI-----VKYRFSWP-----PREP--VTRDEARDY 1381
Db 1250 GKPSKRLHRIETDEESCDNAHGDAQARDSPRLPSPQESTKKPKYRIESDEEDFE 1309
Qy 1382 DV---ITHPMDQTV 1393
Db 1310 NVKVGSPLDYSLV 1323

RESULT 45

US-10-032-585-7646
; Sequence 7646, Application US/10032585
; Publication No. US20030180953A1
; GENERAL INFORMATION:
; APPLICANT: Terry, Roemer D.
; APPLICANT: Bo, Jiang
; APPLICANT: Charles, Boone
; APPLICANT: Howard, Bussey
; TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery
; FILE REFERENCE: 10182-005-999
; CURRENT APPLICATION NUMBER: US/10/032,585
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 8000
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 7646
; LENGTH: 1881
; TYPE: PRT
; ORGANISM: Candida albicans
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (1881)..(1881)
; OTHER INFORMATION: X=any amino acid
US-10-032-585-7646

Query Match 3.7%; Score 293; DB 14; Length 1881;
Best Local Similarity 19.5%; Pred. No. 1.2e-07;
Matches 278; Conservative 218; Mismatches 516; Indels 416; Gaps 59;

Qy 17 PGEPPFFTIPIH-----TQAFRTREEVEARLERYSERIWTCKSTGSSQLTHK 63
Db 717 PSSPFRGKISYELPEELDTKVLNLTKELOTEKNAESNDKELNEKIE--KLTNLSTKLET 774
Qy 64 EAWBEEQVABLKEEPPAWYKLVLEM-----VHNHTASLEKLVDTAWLIMTKYAVGE 118
Db 775 KLEDQEQLAK-IQEDHKSLNEKFLVTANSLOGIKARTKESSETISGPDQQLQEAALKGN 833
Qy 119 ECDPEVGKMKLVKVIKHLPEKVD--SEATEKSDGACDS-----PSSDKENSQI 169
Db 834 TSESTL---KQK-----EKLDSQEAKKLEDGINNMTRDLFHLKKSKEAETQI 881
Qy 170 AQDHQKKEVTVKDEGRRE-----SINDRARSPRKLPSTSLKKGKRWAPPKPLPHY 222
Db 882 KQREPKNLYTEFNTKYKQYELQINNLNKNSEFKQKINELSKKIESLTEDNKENAKOL 941
Qy 223 DVKLQ-----NE-----DKIIS-NVPADSLIRTERPNKEIVRYFIRHNALRAGTG--ENAP 271
Db 942 EEKLRDTENNHEHLMKLRSAVAYNDLKAKASEEETVK---AKELETLTSTIDNLE 998
Qy 272 WVVEDELVKYSLSKSDFLDPYKYWTLPNPSYKRNKNTGSPDR-KPSKSKKTDSNLSLS 330
Db 999 KELKEQSKQKVEGLQGN-----ITDSNTEKFKLEDELKSIKKNKEITSQNS 1048
Qy 331 PLNPKLWCHVHLKKSLSGSPKVKNSKNSKSPBEHLEEMKMGSPNKLHTTNPFIKPKGPP 390

Db 1271 GERARAEINDKVHKLQN-EVESVTG-----MLNEAEGKAIFAKLAKDVASLSSQLQDTQE 1322
Qy 623 LLLPDAQPIPTAVSLMEALSADKGGFLYNRVLLQTLTLLQDEIAED----- 674
Db 1323 LLQETROKLVSTKRLQLEEBRNS-----LQDQDEMEAKQNL 1363
Qy 675 --YGLGKMLSEIPLTLHVSSELVR-LCLRRSDVOESEGSDTDNDKDSAAPE-----DN 726
Db 1364 RHISTNLQSDSKKKLQDFASTVBALEGGKRFQKIEIENLTQOYEEKAAAYDKLEKYN 1423
Qy 727 EVQDEF-----LEKLETSEFFELTSEBKQLIALTALCHRLMVTYQVQDHM 770
Db 1424 RLQOEELDLVDLQNRQLVSNLEK-KQKFDQLAAEEK-----NISKYA----- 1468
Qy 771 ETRQMSAEL-WKERLAV-----LKEENDKRAEKQKRAKMEAKNKGKVENGLG 820
Db 1469 DERDRAEAAREKETKALSARALEEALEAELEERTNKLKAEMEDLVSSKDDVGKXNVH 1528
Qy 821 KTDKRRIVKPEPOVD-----TEADMISAVKSRL-----LAIQAKKEREIOER--- 865
Db 1529 ELEKSKRAL--ETQMEEMKTQLEEDLEQATDAKLRLVNMQALKQOFERDLQARDEQ 1586
Qy 866 --EMKVKLQ-----AEEERIRKHAARAFQEGIAKAKLVMRRTPI-GTDRNHR 915
Db 1587 NEEKRQRLQRLHEYTELEDERKQALAAAKKLEGDLKOLEQADSAIKGRE----- 1641
Qy 916 YWLFDEVPGLPIEGWVHSDIDYRNNHCKDHTVSGDEDYC---PRSKANLGNKASMN 972
Db 1642 ----EAIKQLRKLQAKMKD-----FORELEDARASRDEIFATAKENEKKALEADLMQ 1691
Qy 973 TOHGAT-----EVAVETTPKQONLWFLCDSQKQELDELLNCLHPQIGRES 1019
Db 1692 LQEDLAAERARKQADLEKEELAEELASSLSGRNA--LQDEKRRLE-----ARIA 1739
Qy 1020 OLKERLEKRYQDIHSIHLARKPNLGLKSCDQGNQELLNFLRSDLTIEVATRLQKGLGYVE 1079
Db 1740 QLEELEEBE-----QGNMEAM-----SDRYRKATQ----- 1764
Qy 1080 ETSEFEARVISLEKLFQFECVIALQASVIKFLQGFMAPOKRRKLOSEDSAKTEEVD 1139
Db 1765 ----QAEQLSNE-----LATERSTAQ-----NESARQOLERQNKELASKLHE 1803
Qy 1140 EKKMVE-----EAKVASALEKWKATAREAOPTFSRMHVLGLMGLDCAIKWDSASNA 1189
Db 1804 MEGAVSKFSTIAALEAKIAQLEQVQOEAKEQAATK----- 1842
Qy 1190 RCKVCPKKGEDKILCDECNKAFHLCIRPALYEVDPGEMQCPACQATARNRGRNY 1249
Db 1843 ----SLQKQKK-----LKEILLQVEDERKMAEQYKQAEKGNARVKOL 1882
Qy 1250 TB--ESASEDSDESDSEEBEEDYEVAGLRPRPKT--IRGKHSVIPPAARS 1305
Db 1883 KQLEAEAEESORINARRKQRLQRELDATESNEAMGREVNALKSLRGRNETSFVPSRRS 1942
Qy 1306 GBR 1308
Db 1943 GBR 1945

RESULT 47

US-10-341-434-103
; Sequence 103, Application US/10341434
; Publication No. US20030215835A1
; GENERAL INFORMATION:
; APPLICANT: Origene Technologies
; TITLE OF INVENTION: Differentially Regulated Prostate Cancer Genes
; FILE REFERENCE: 9U 204 205 R1
; CURRENT APPLICATION NUMBER: US/10/341,434
; PRIOR FILING DATE: 2003-07-18
; PRIOR APPLICATION NUMBER: US 60/348,164
; PRIOR FILING DATE: 2002-01-15
; PRIOR APPLICATION NUMBER: US 60/348,119
; PRIOR FILING DATE: 2002-01-15

Query Match 3.6%, Score 290.5, DB 15; Length 1972;
Best Local Similarity 19.9%; Pred. No. 1.7e-07;
Matches 275; Conservative 214; Mismatches 507; Indels 387; Gaps 58;
Qy 75 LLKEEPPAWYKEL-VLEMVHNTASLEKLVDTAMLEIMTKY-----AVGEEDC 121
Db 801 LARKAFARQOQLTAMKVIQNCAYLKLNRWWRFLTKVKKPLLOVTRQBEEMQAKDE 860
Qy 122 FEVGKEMKLVKIKIHPLEKVDBEATEKKS-----DGACDSSPSDKENSQIAQDHQK 175
Db 861 LQTKERQKAE-NELKELEQKHSQLTTEKLLQEQLAETELYAAEEMRVLAARKQE 919
Qy 176 KETVVEDEGRRESINDRARRSPKLPSTLSKKGGRKMAPPLPHKYDVKLQNEKIIISN 235
Db 920 LEEILHEMARLEEBEEDRGQ-----LQABRK-----KMAOQMLDLEEQLEB--- 961
Qy 236 VPADSLIRTPPNKEIVRYEIRHNALRAGTGENAPVVEDELVKKYSLSPKFS---DFL 292
Db 962 -----EEAARQKLQ---LEKVTABAKIKKEDEILVMDQNNKLSKERKLL 1004
Qy 293 LDPKYMTLPNTSKRKTGSPDRKPSKSKTDNSLSPLNPKLVCHVHLKKSLSG--- 348
Db 1005 EERISDLTTLNLAEEBKA-----KNLTKLKNKHESMISEL-----EVLAKKEERSQEL 1053
Qy 349 SPLKVNKSNKSPPEHLEEMKMSPNKLTNTHPIPKGPPAKPGKHGSKPLKAKGRS 408
Db 1054 EKLKRLKLGSDAFHEQIADLQAOIAELM-----QLAKKBELOALARLDDEIAQKNA 1109
Qy 409 -KGI-----LNGOKSTGNSKSPK-----GLKTPKTKMKTLLDMAKGTQK 449
Db 1110 LKKTRELGHTSDLOEDLDSERAARKAEKQRRDLGSELEAKLTELED--TLD-STATQ 1166
Qy 450 MTRAPRNSGGT-----PRTSSKPHKLPAPALHLIAYYKENKDKRKSALSCVISKT 502
Db 1167 ELRAKREQEVTVLKALDEETRSHEAOVQEMRQKHAQAVEELTEQLEQFKRA-KANLDKN 1225
Qy 503 ARLLSSDEBRAPLPELRSLOKRYELLEHKKRWASMESEORKEYLKKRBEELKKLKEKA 562
Db 1226 KQTLKEN-ADLAGESELVLGAKQE-VEHKKK-----KLEAQVQELQSKCSD-- 1270
Qy 563 KERREKEMLEKQKRYEDQELTGKNLPAFRLVDTPEGLPNTLFGDVAMVVEFLSCYSG 622
Db 1271 GERARAEINDKVHKLQN-EVESVTG-----MLNEAEGKAIFAKLAKDVASLSSQLQDTQE 1322
Qy 623 LLLPDAQPIPTAVSLMEALSADKGGFLYNRVLLQTLTLLQDEIAED----- 674
Db 1323 LLQETROKLVSTKRLQLEEBRNS-----LQDQDEMEAKQNL 1363
Qy 675 --YGLGKMLSEIPLTLHVSSELVR-LCLRRSDVOESEGSDTDNDKDSAAPE-----DN 726
Db 1364 RHISTNLQSDSKKKLQDFASTVBALEGGKRFQKIEIENLTQOYEEKAAAYDKLEKYN 1423
Qy 727 EVQDEF-----LEKLETSEFFELTSEBKQLIALTALCHRLMVTYQVQDHM 770
Db 1424 RLQOEELDLVDLQNRQLVSNLEK-KQKFDQLAAEEK-----NISKYA----- 1468
Qy 771 ETRQMSAEL-WKERLAV-----LKEENDKRAEKQKRAKMEAKNKGKVENGLG 820
Db 1469 DERDRAEAAREKETKALSARALEEALEAELEERTNKLKAEMEDLVSSKDDVGKXNVH 1528
Qy 821 KTDKRRIVKPEPOVD-----TEADMISAVKSRL-----LAIQAKKEREIOER--- 865
Db 1529 ELEKSKRAL--ETQMEEMKTQLEEDLEQATDAKLRLVNMQALKQOFERDLQARDEQ 1586
Qy 866 --EMKVKLQ-----AEEERIRKHAARAFQEGIAKAKLVMRRTPI-GTDRNHR 915

```
Db 1587 NEEKROLOQLHVEYETELEDERKORALAAAKKLEGLDKOLELQADSAIKGRE----- 1641
QY 916 YWLFSDVEVPGLFTEKGMVHDSIDYRFNHCKDHTVSGDEDYC---PRSKANLGNKASWN 972
Db 1642 -----BAIKOLRLKLOQMKD-----FORELEDAASRDEIFATAKENEKAKSLEADLMQ 1691
QY 973 TQHGAT-----EVAVETTPKQGNLWFLCDSOKELDELNCLHPQGIRES 1019
Db 1692 LQEDLAAARARQADLEKEBELAELASSLSGRNA--LQDEKRRLE-----ARIA 1739
QY 1020 OLKERLEKRYQDIHSHILARKNLGLKSCDGNQELLNFRSLIEVATLQXGGLGYVE 1079
Db 1740 QLEEELEEE-----QGNMEAM-----SDRVKATQ----- 1764
QY 1080 ETSEFEARVISLEKLDGFCGVIALQASVTKFLQGMAPKQKRRKQSEDSAKTEEVDE 1139
Db 1765 -----QAEQLSNE-----LATERSTAQK-----NESARQQLERQNKELRSKLHE 1803
QY 1140 EKQWVE-----EAKVASALEKWKTAIREAQTFSRMHVLLGLMDACIKWDMSAENA 1189
Db 1804 MEGAVSKFKSTTAALEAKIAQLEEQVEQEAEREKQAATK----- 1842
QY 1190 RCKVCPKSGEDDKLILCDECNKAFHFLCLRPALYEVDPGEWQCPACQATARNRGRNY 1249
Db 1843 -----SLKQDKK-----LKEILLOVEDERKMAEQYKEQAEGNARVKOL 1882
QY 1250 TE--ESASDSEDSDEDEEBEEREBEEDYEVAGRLRPRKT--IRGKHSVIPPAAKS 1305
Db 1883 KRQLEAEESQRINANRRKRLQRELDATESNEMGRENVALSKURRNETSFVPSRRS 1942
QY 1306 GRR 1308
Db 1943 GGR 1945
```

RESULT 48

```
US-09-927-597-4
; Sequence 4, Application US/09927597
; Publication No. US20030032018A1
; GENERAL INFORMATION:
; APPLICANT: Malik, Pady
; APPLICANT: Beraud, Christophe
; APPLICANT: Freedman, Richard
; APPLICANT: Craven, Andrew
; APPLICANT: Sakowicz, Roman
; APPLICANT: Hartman, James
; TITLE OF INVENTION: Human smooth muscle myosin heavy chain
; FILE REFERENCE: CYTOF018
; CURRENT FILING DATE: 2001-08-10
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 1979
; TYPE: PRT
; ORGANISM: Human
US-09-927-597-4
```

```
Query Match 3.6%; Score 290.5; DB 10; Length 1979;
Best Local Similarity 19.9%; Pred. No. 1.8e-07;
Matches 275; Conservative 214; Mismatches 507; Indels 387; Gaps 58;

QY 75 LKKEEFPANVEKL-VLEMVHNTASLEKLVDTAWLIMTKY-----AVGECD 121
Db 808 LARKAFKQOQLTAMTQORNAAYKLKRNQMWRFLTKVKPPLQVTRQESMQAKDE 867
QY 122 FEVGKEMLKVKIVKIHLEKVDDEATEKKS-----DGACDPSDDKENSIOAQDHOK 175
Db 868 LQTKERQQAEB-NELKELEKHSQLTTEKQLLQEQLOAQETELYAEAEEMRVLAQKQE 926
QY 176 KETVVKDEGRRESINDRARRSPKLPSTSLKGERKWAPPKFLPHKYDVKLQNEDKIISN 235
```

```
Db 927 LBEILHEMPEARLEEBEDRQQ-----LOAERK-----KWAQQMLDLEBQLEE----- 968
QY 236 VPADSLIRTERPKNKEIVRYFIRHNALRAGTGENAPWVVEDELVKKYSILSPKFS-----DPL 292
Db 969 -----SEARQKLQ---LEKVTAEAKIKLEDEILVMDQNNKLSKERKUL 1011
QY 293 LPYKYMTLNPSTKRNKNTGSPDRKSKKGTNNSSSPINPKLWCHVHLKKSLSG----- 348
Db 1012 EERISDLATTLNABEBA-----KNLTKLKNKHESMISEL-----EVLKKEEKSRQEL 1060
QY 349 SPLKVNKSNKSPHEHLEEMKMWSPNKLHTNFHPKGPAPKPKGKSHDKPLKAKGRS 408
Db 1061 EKLKRRKLEGDSDFHEQIADLQQAELAKN-----QLAKKEEELQAAALRDLDEIAOKNNA 1116
QY 409 -KGI-----LNGQKSTGNSKSPK-----GLTKPTKMKQMTLTDMAKGTQK 449
Db 1117 LKKIRELEGHISLQEDLDSERAARNKAEKQKRDLEGELEALKTELED--TLD-STATQQ 1173
QY 450 MTRAPNNSGGT-----PRTSSKPHKHPAALHLIAYYKKNKOREDKRSALSVCISKT 502
Db 1174 ELRAKREQVTVLKKALDEBTRSHQAQVQEMRQKHAQAVEELTEQLEQFKRA--KANLDKN 1232
QY 503 ARLLSSEDEARLPEELRSIVOKRYELHLEHKKERWASSEORKEYLKKKEELKKLKEKA 562
Db 1233 KOTLEKEN--ADLAGELRVLGQAQE--VEHKKK-----KUEAQVOELQSKCSD-- 1277
QY 563 KERREKEMLERLEKOKRYEDOBELTGKNLPAPFLVDTPTEGLPNTLFGDVAMVVEFFLSCYS 622
Db 1278 GERAAELNDKVHKLQN--EVESVTG-----MLNEAEGKAIKLAKDVASLSLQLODTQE 1329
QY 623 LLLPDAQVPITAVSMEALSADKGFYLNRLVILLQTLLOTLQDEIAED----- 674
Db 1330 LLQEEETRQKLVNSTKLRLQLEERN-----LQDQDDEMEAKONLE 1370
QY 675 --YGEKMKLSEIPLTLHSVSELVR--LCLLRSSDVQEESEGSTDDNKKDSAAFE-----DN 726
Db 1371 RHISTNLQLSDSKKKLODFASTVBALEEGKRFQKEIENLTQQYEEKAAAYDKLEKTKN 1430
QY 727 EVQDEF-----LEKLTSEFFELTSEEKLOILTALCHRLTILMTYSVQDHM 770
Db 1431 RLQOELDDLVDLDNQRLVSNLEK--KQKFPOLLAEK-----NISSKYA----- 1475
QY 771 ETRQQMSAEL-WKERLAV-----LKEENDKRAEKQKKEKEMAKKNKENGKENGVLG 820
Db 1476 DERDRAEAEREKETKALSARALEEALEKELEERTNKMKAEMEDLVSSKDDVGVKNVH 1535
QY 821 KYDRKRVKFPQVD-----TEADMISAVKSRRL-----LAQAKKEREIOER--- 865
Db 1536 ELEKSKRAL--ETQMEEMKTQLESELELOQATEDAKLRLEVNMQALKGOFERDLQARDEQ 1593
QY 866 --EMKVTLERQ-----AEEERIRKHAAAEKAFQEGIAKAKLVNRRTPI--GTDNRHNR 915
Db 1594 NEEKRQRLQRLHVEYTELEDERKQALANAANKKLEGLDKOLELQADSAIKGRE----- 1648
QY 916 YWLFSDVEVPGLFIEKGMVHDSIDYRFNHCKDHTVSGDEDYC---PRSKANLGNKASWN 972
Db 1649 -----EAIKQLAKLQAMKD-----FORELEDAASRDSIFATAKENEKAKSLEADLMQ 1698
QY 973 TQHGAT-----EVAVETTPKQGNLWFLCDSOKELDELNCLHPQGIRES 1019
Db 1699 LQEDLAAARARQADLEKEBELAELASSLSGRNA--LQDEKRRLE-----ARIA 1746
QY 1020 OLKERLEKRYQDIHSHILARKPNLGLKSCDGNQELLNFRSLIEVATRLQKGGGLGYVE 1079
Db 1747 QLEEELEEE-----QGNMEAM-----SDRVKATQ----- 1771
QY 1080 ETSEFEARVISLEKLDGFCGVIALQASVTKFLQGMAPKQKRRKQSEDSAKTEEVDE 1139
Db 1772 -----QAEQLSNE-----LATERSTAQK-----NESARQQLERQNKELRSKLHE 1810
QY 1140 EKQWVE-----EAKVASALEKWKTAIREAQTFSRMHVLLGLMDACIKWDMSAENA 1189
Db 1811 MEGAVSKFKSTTAALEAKIAQLEEQVEQEAEREKQAATK----- 1849
```


Qy	1479	QSRDEDRRSREAEIQ	1493
		: : :	
Dd	1447	LVGVDKSKHRDLEEQ	1461

RESULT 50

```

US-10-029-386-30379
; Sequence 30379, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David K.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
; FILE REFERENCE: AEOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34286
; SOFTWARE: Annotmax Sequence Listing Engine vers. 1.1
; SEQ ID NO 30379
; LENGTH: 58
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO CHR7.1
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.9
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.86
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.8
; OTHER INFORMATION: SWISSPROT HIT: Q9JLNS, EVALUAE 4.70e+00
US-10-029-386-30379

```

```
Query Match          3.6%; Score 284; DB 14; Length 58;
Best Local Similarity 100.0%; Pred. No. 5.2e-09;
Matches 56; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

Qy	1033	IHSIHLAKPNLGLKSCDGNQELLNFRSLDI EVATRLQKGIGLYVEETSEFARV	1088
Dd	1	IHSIHLAKPNLGLKSCDGNQELLNFRSLDI EVATRLQKGIGLYVEETSEFARV	56

Search completed: March 9, 2005, 15:11:59
Job time : 168.292 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 9, 2005, 14:24:09 ; Search time 133.045 Seconds
(without alignments)
4877.138 Million cell updates/sec

Title: US-10-702-148-21
Perfect score: 10138
Sequence: 1 MGQTKSTSGGNKCNQEQ.....AGHNRKYEKKWTDFKVS 1972

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1391452 seqs, 329044822 residues

Total number of hits satisfying chosen parameters: 1391452

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 65 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/PTCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/PTCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/1/pubpaa/US10D_PUBCOMB.pep.*
- 17: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pep.*
- 19: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
- 20: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	10138	100.0	1972	9	US-09-839-479-21
2	10138	100.0	1972	15	US-10-376-537-21
3	10138	100.0	1972	15	US-10-702-148-21
4	10022.5	98.9	1969	9	US-09-839-479-71
5	10022.5	98.9	1969	15	US-10-376-537-72
6	10022.5	98.9	1969	15	US-10-702-148-71
7	2512	24.8	1873	13	US-10-087-192-666
8	2493.5	24.6	1876	9	US-09-839-479-70
9	2493.5	24.6	1876	15	US-10-376-537-71
10	2493.5	24.6	1876	15	US-10-702-148-70
11	2482.5	24.5	1878	9	US-09-839-479-13
12	2482.5	24.5	1878	15	US-10-376-537-13
13	2482.5	24.5	1878	15	US-10-702-148-13

ALIGNMENTS

RESULT 1

US-09-839-479-21
; Sequence 21, Application US/09839479
; Publication No. US2002003979A1
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042002
; CURRENT APPLICATION NUMBER: US/09/839,479
; PRIOR FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24

14	2300	22.7	1586	13	US-10-087-192-663	Sequence 663, Appl
15	1021	10.1	194	9	US-09-864-761-33808	Sequence 33808, A
16	888	8.8	176	9	US-09-839-479-65	Sequence 65, Appl
17	888	8.8	176	15	US-10-376-537-66	Sequence 66, Appl
18	888	8.8	176	15	US-10-702-148-65	Sequence 65, Appl
19	817	8.1	160	9	US-09-764-864-1185	Sequence 1185, Ap
20	726.5	7.2	1674	9	US-09-839-479-61	Sequence 1, Appli
21	726.5	7.2	1674	15	US-10-376-537-1	Sequence 1, Appli
22	726.5	7.2	1674	15	US-10-702-148-1	Sequence 69, Appl
23	709	7.0	1673	9	US-09-839-479-69	Sequence 69, Appl
24	709	7.0	1673	15	US-10-376-537-70	Sequence 70, Appl
25	709	7.0	1673	15	US-10-702-148-69	Sequence 69, Appl
26	582	5.7	1334	9	US-09-764-864-1578	Sequence 1578, Ap
27	538.5	5.3	1525	9	US-09-839-479-68	Sequence 68, Appl
28	538.5	5.3	1525	15	US-10-376-537-69	Sequence 69, Appl
29	538.5	5.3	1525	15	US-10-702-148-68	Sequence 68, Appl
30	533	5.3	1531	9	US-09-839-479-29	Sequence 29, Appl
31	533	5.3	1531	15	US-10-376-537-29	Sequence 29, Appl
32	533	5.3	1531	15	US-10-702-148-29	Sequence 29, Appl
33	532	5.2	1527	9	US-09-839-479-27	Sequence 27, Appl
34	532	5.2	1527	15	US-10-376-537-27	Sequence 27, Appl
35	532	5.2	1527	15	US-10-702-148-27	Sequence 27, Appl
36	491	4.8	97	9	US-09-864-761-37847	Sequence 37847, A
37	463	4.6	91	9	US-09-864-761-36754	Sequence 36754, A
38	445.5	4.4	175	9	US-09-839-479-64	Sequence 64, Appl
39	445.5	4.4	175	15	US-10-376-537-65	Sequence 65, Appl
40	445.5	4.4	175	15	US-10-702-148-64	Sequence 64, Appl
41	406	4.0	196	14	US-10-062-831-94	Sequence 94, Appl
42	406	4.0	196	14	US-10-062-599-94	Sequence 94, Appl
43	390	3.8	2781	15	US-10-263-929-122	Sequence 122, App
44	372.5	3.7	3225	16	US-10-408-765A-254	Sequence 254, App
45	363	3.6	70	9	US-09-864-761-33806	Sequence 33806, A
46	357	3.5	1564	15	US-10-144-198-2	Sequence 2, Appli
47	357	3.5	1564	15	US-10-144-198-4	Sequence 4, Appli
48	326.5	3.2	1162	11	US-09-894-273-2	Sequence 2, Appli
49	326.5	3.2	1162	14	US-10-294-804-2	Sequence 2, Appli
50	326.5	3.2	2062	15	US-10-052-648A-52	Sequence 52, Appl
51	325.5	3.2	2560	15	US-10-263-929-142	Sequence 142, App
52	322	3.2	2665	9	US-09-864-761-34248	Sequence 34248, A
53	322	3.2	3664	14	US-10-177-293-423	Sequence 423, App
54	322	3.2	3664	15	US-10-263-929-143	Sequence 143, App
55	322	3.2	3664	16	US-10-408-765A-2287	Sequence 2287, Ap
56	320	3.2	2375	16	US-10-408-765A-277	Sequence 277, App
57	318.5	3.1	2492	16	US-10-697-526-2	Sequence 2, Appli
58	316	3.1	1798	10	US-09-981-151A-48	Sequence 48, Appl
59	315.5	3.1	3051	15	US-10-144-194A-62	Sequence 62, Appl
60	310	3.1	1875	15	US-10-369-493-22285	Sequence 22285, A
61	307	3.0	2844	15	US-10-267-502-370	Sequence 370, App
62	305.5	3.0	2843	8	US-08-681-219-32	Sequence 32, Appl
63	305.5	3.0	2843	9	US-09-987-482-1	Sequence 1, Appli
64	305.5	3.0	2843	10	US-09-230-111C-30	Sequence 30, Appl
65	305.5	3.0	2843	14	US-10-092-138-30	Sequence 30, Appl

! PRIOR APPLICATION NUMBER: JP 9/116570
! PRIOR FILING DATE: 1997-04-18
! NUMBER OF SEQ ID NOS: 72
! SOFTWARE: FastSeq for Windows Version 4.0
! SEQ ID NO 21
! LENGTH: 1972
! TYPE: PRT
! ORGANISM: Homo sapiens
US-09-839-479-21

Query Match 100.0%; Score 10138; DB 9; Length 1972;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1972; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MGQKSTSSGGNRKCNQEQSKNQPLDARVDKI	KDKKPRKKAMSSSSNSDSGTS	SDTS	60
Db	1	MGQKSTSSGGNRKCNQEQSKNQPLDARVDKI	KDKKPRKKAMSSSSNSDSGTS	SDTS	60
Qy	61	SEGISSSDSDLEDEDEEDOSIESEDDDSSE	EAQHKNNOVLLHGISDPKADGKA		120
Db	61	SEGISSSDSDLEDEDEEDOSIESEDDDSSE	EAQHKNNOVLLHGISDPKADGKA		120
Qy	121	TEKAQEKRIHQPLAFESQTHSFOSQKQPVLS	QQLPFIQSSQAKESVNVKHTSVIQ		180
Db	121	TEKAQEKRIHQPLAFESQTHSFOSQKQPVLS	QQLPFIQSSQAKESVNVKHTSVIQ		180
Qy	181	STGLVSNVKPLSLVNOAKETMYKLI	VPSDVLKAGNKTSESSLLTSELRSKREQYKQ		240
Db	181	STGLVSNVKPLSLVNOAKETMYKLI	VPSDVLKAGNKTSESSLLTSELRSKREQYKQ		240
Qy	241	APPSQLKQESSKSLKVAIALSNPKATSS	SPAPHPKOTLENNHPNPFITNALLGNHPNG		300
Db	241	APPSQLKQESSKSLKVAIALSNPKATSS	SPAPHPKOTLENNHPNPFITNALLGNHPNG		300
Qy	301	VIQSVIOBAPALATTKTKQKQKINENIAA	ASTPSSPVNLTSGRRTPGQTPVMPGAS		360
Db	301	VIQSVIOBAPALATTKTKQKQKINENIAA	ASTPSSPVNLTSGRRTPGQTPVMPGAS		360
Qy	361	PILHSQGEKAVSNVNVKTKQHSHPAKSL	VEQPRGTDSDIPSSKQSEDSNEDEEDE		420
Db	361	PILHSQGEKAVSNVNVKTKQHSHPAKSL	VEQPRGTDSDIPSSKQSEDSNEDEEDE		420
Qy	421	EDEDEDDSDSDSQSESDNSSESDT	BGSEEDDDDKQDESDDTGEKTSMKLNKTT		480
Db	421	EDEDEDDSDSDSQSESDNSSESDT	BGSEEDDDDKQDESDDTGEKTSMKLNKTT		480
Qy	481	SSVKSPPGMSLTGHSTPRNLHIAPGSA	PAALCSSESQSPALFGTSSSTLTSSPHSGTSKR		540
Db	481	SSVKSPPGMSLTGHSTPRNLHIAPGSA	PAALCSSESQSPALFGTSSSTLTSSPHSGTSKR		540
Qy	541	RVVTDERELRIPLVYQWQRETRIRNFG	GRLOGEVAYYAPCGKLRQYPEVITYLSRNGIM		600
Db	541	RVVTDERELRIPLVYQWQRETRIRNFG	GRLOGEVAYYAPCGKLRQYPEVITYLSRNGIM		600
Qy	601	DISRDNFSFAKIRVGDYFYEARDGQEN	QWCLLKEEDVIPRIRAMEGRGRPPNPDRORA		660
Db	601	DISRDNFSFAKIRVGDYFYEARDGQEN	QWCLLKEEDVIPRIRAMEGRGRPPNPDRORA		660
Qy	661	FEESRMRRKGRPNVGNNAEFLDNADAK	LRLKQAEIARQAAQIKLRLKQKQEQARVA		720
Db	661	FEESRMRRKGRPNVGNNAEFLDNADAK	LRLKQAEIARQAAQIKLRLKQKQEQARVA		720
Qy	721	KEAKKQQAIAAAEKRKQEQIKIMQOQEK	IKR100IRMEKELRAQOILEAKKKKEEAA		780
Db	721	KEAKKQQAIAAAEKRKQEQIKIMQOQEK	IKR100IRMEKELRAQOILEAKKKKEEAA		780
Qy	781	NAKLLAEAKRIKEKEMRQQAQVLLKH	QERRRRQHMMLKAMAEARKKAEKERLKQEKRD		840
Db	781	NAKLLAEAKRIKEKEMRQQAQVLLKH	QERRRRQHMMLKAMAEARKKAEKERLKQEKRD		840
Qy	841	EKRLNKERKLQORLEMAKELKPNEDM	CLADQKPLPELPRIPGLVLSGTSFSDCLMW		900
Db	841	EKRLNKERKLQORLEMAKELKPNEDM	CLADQKPLPELPRIPGLVLSGTSFSDCLMW		900

Qy	901	VOFLRNFGKVLGPDVNDVPNLSVLQEG	LLNIGDSMGVQDILLVRLLSAAVCDPGLITGY		960
Db	901	VOFLRNFGKVLGPDVNDVPNLSVLQEG	LLNIGDSMGVQDILLVRLLSAAVCDPGLITGY		960
Qy	961	KAKTALAGEHLLNVGNRDNVSEILQIF	MEARCGOTELTESLKTKAFAQHTTQA	KASVLAF	1020
Db	961	KAKTALAGEHLLNVGNRDNVSEILQIF	MEARCGOTELTESLKTKAFAQHTTQA	KASVLAF	1020
Qy	1021	LINELACSKSVVSEIDKNIDYMSNLR	DRDKWVVEGKLRLRIIHAKTGKROTSG	IDIGE	1080
Db	1021	LINELACSKSVVSEIDKNIDYMSNLR	DRDKWVVEGKLRLRIIHAKTGKROTSG	IDIGE	1080
Qy	1081	EQHPLGTPTPKRRRRKGGSDYDDDD	DDDDDDDEDEDEDEKDGKKTDCI	CEBDD	1140
Db	1081	EQHPLGTPTPKRRRRKGGSDYDDDD	DDDDDDDEDEDEDEKDGKKTDCI	CEBDD	1140
Qy	1141	EGDQAASVEELEKQIEKLSKQSQYRR	KLPDASHLSRVFMPDRYRRYRWILPR	CGGIF	1200
Db	1141	EGDQAASVEELEKQIEKLSKQSQYRR	KLPDASHLSRVFMPDRYRRYRWILPR	CGGIF	1200
Qy	1201	VEGMESEGLEBEIAKERREKLKAES	VQIKEMFETSGDSLNCSTNDHCEQED	LKEKONT	1260
Db	1201	VEGMESEGLEBEIAKERREKLKAES	VQIKEMFETSGDSLNCSTNDHCEQED	LKEKONT	1260
Qy	1261	NLFLOKPGSFSKLSKLEVAKMPPESE	VMTPKPNAGANGCTLSYQNSCKHSL	GSVQSTAT	1320
Db	1261	NLFLOKPGSFSKLSKLEVAKMPPESE	VMTPKPNAGANGCTLSYQNSCKHSL	GSVQSTAT	1320
Qy	1321	QSNVEKADSNLNFNTGSSGPGKFY	SPLPNDQLLTKLTTEKNQWFSLLPRT	PCDDTSLTHA	1380
Db	1321	QSNVEKADSNLNFNTGSSGPGKFY	SPLPNDQLLTKLTTEKNQWFSLLPRT	PCDDTSLTHA	1380
Qy	1381	DMSTASLVTPOSQPPSKSPPTPAP	LGSSAQNPGVLPALSPLOVKGVMGLO	FCGW	1440
Db	1381	DMSTASLVTPOSQPPSKSPPTPAP	LGSSAQNPGVLPALSPLOVKGVMGLO	FCGW	1440
Qy	1441	PTGVVTNIPFTLSVPSIGSLGSEG	NGNSFNTSNVASSKSESPVQNEKATS	AQPAAV	1500
Db	1441	PTGVVTNIPFTLSVPSIGSLGSEG	NGNSFNTSNVASSKSESPVQNEKATS	AQPAAV	1500
Qy	1501	EVAKPVPFPSPKPIPEEMQFCWRI	IDPEDLKALLKVLHLRGIREKALQKQ	IQKHLDYIT	1560
Db	1501	EVAKPVPFPSPKPIPEEMQFCWRI	IDPEDLKALLKVLHLRGIREKALQKQ	IQKHLDYIT	1560
Qy	1561	OACLKNKDVAIIELNENEENQVTRD	I VENWSVEQAMEMDLSVLOQVEDL	ERRVASASLQ	1620
Db	1561	OACLKNKDVAIIELNENEENQVTRD	I VENWSVEQAMEMDLSVLOQVEDL	ERRVASASLQ	1620
Qy	1621	VKGWMCPEPASERDLVYFHKSTFK	LCKBHDGFTGEDESSAHALERKSNP	LIDIAVTR	1680
Db	1621	VKGWMCPEPASERDLVYFHKSTFK	LCKBHDGFTGEDESSAHALERKSNP	LIDIAVTR	1680
Qy	1681	LADLERNIERIEEDIAFGLVRRRAL	SEARSAQVALCIQOLQKSI	AWEKSIIMKYCOI	1740
Db	1681	LADLERNIERIEEDIAFGLVRRRAL	SEARSAQVALCIQOLQKSI	AWEKSIIMKYCOI	1740
Qy	1741	CRKGDNEBELLLLCGCDKGCHTY	CHRPKIITIPDGDWFCPACIAXASQ	TLIKKLHVKG	1800
Db	1741	CRKGDNEBELLLLCGCDKGCHTY	CHRPKIITIPDGDWFCPACIAXASQ	TLIKKLHVKG	1800
Qy	1801	KKTNESKKGKVTITGDTDEDEDS	ASTSSSLKRGNKDLQKRKWEENTS	INLSKQESFTSVK	1860
Db	1801	KKTNESKKGKVTITGDTDEDEDS	ASTSSSLKRGNKDLQKRKWEENTS	INLSKQESFTSVK	1860
Qy	1861	KPKRDDSXDALCALCSMILTEMETH	EDAWPFLPLVNLKLVPGYKVKIK	KKPMDFSTIREKLSS	1920
Db	1861	KPKRDDSXDALCALCSMILTEMETH	EDAWPFLPLVNLKLVPGYKVKIK	KKPMDFSTIREKLSS	1920
Qy	1921	GOYNLETTFALDVLVDFNDCET	FPNEDSDDIGRAGHNMRYKFEKK	WDTDFKVS	1972
Db	1921	GOYNLETTFALDVLVDFNDCET	FPNEDSDDIGRAGHNMRYKFEKK	WDTDFKVS	1972

RESULT 2

```

US-10-376-537-21
; Sequence 21, Application US/10376537
; Publication NO. US2003022405A1
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042001
; CURRENT APPLICATION NUMBER: US/10/376,537
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US/09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 21
; LENGTH: 1972
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-376-537-21

```

Query Match	100.0%	Score 10138;	DB 15;	Length 1972;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 1972;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	MGQTKTSSGGGNKCKQEQSKNQPLDARVDKIKDKPKRKAMSSSSDSDSGTSSDTS	60	
Db	1	MGQTKTSSGGGNKCKQEQSKNQPLDARVDKIKDKPKRKAMSSSSDSDSGTSSDTS	60	
Qy	61	SEGISSSDSDLEDEEEDQSIIESEDDDDSDSESAQHKSNNOVLLHGISDPKADGOKA	120	
Db	61	SEGISSSDSDLEDEEEDQSIIESEDDDDSDSESAQHKSNNOVLLHGISDPKADGOKA	120	
Qy	121	TEKAQEKRIHQPLPLAFESQTHSFQSQOKQOPVLSQQLPFFQSQSAKEEVSNNKHTSVIQ	180	
Db	121	TEKAQEKRIHQPLPLAFESQTHSFQSQOKQOPVLSQQLPFFQSQSAKEEVSNNKHTSVIQ	180	
Qy	181	STGLVSNVKPLSLVNQAKKETYMKLIVPSPDVLKAGNKNTEESLLTSELRSKREQVKQ	240	
Db	181	STGLVSNVKPLSLVNQAKKETYMKLIVPSPDVLKAGNKNTEESLLTSELRSKREQVKQ	240	
Qy	241	APFSQLKQKQESSKSLKVIQAALSNPKATSSSPAHPKQTLNHNHPNPFITNALLGNHPNG	300	
Db	241	APFSQLKQKQESSKSLKVIQAALSNPKATSSSPAHPKQTLNHNHPNPFITNALLGNHPNG	300	
Qy	301	VIOSVIOEAPLALTTKTKMQSKINENTAAASSTFFSPVPNLSTGGRTPGQTPVMPSPAS	360	
Db	301	VIOSVIOEAPLALTTKTKMQSKINENTAAASSTFFSPVPNLSTGGRTPGQTPVMPSPAS	360	
Qy	361	PIIHSQKQEKAVSNVNPVKTQHHSHPAKSLVEQFRGTDSDI PSSKQSEDSNEDEEEDDE	420	
Db	361	PIIHSQKQEKAVSNVNPVKTQHHSHPAKSLVEQFRGTDSDI PSSKQSEDSNEDEEEDDE	420	
Qy	421	EEDEEEDDEDSDSQSEDSNSSESDTEGSEEDDDKQDQESDSDTEGKTSMKLNKTT	480	
Db	421	EEDEEEDDEDSDSQSEDSNSSESDTEGSEEDDDKQDQESDSDTEGKTSMKLNKTT	480	
Qy	481	SSVKSPSMSTLGHSTPRNLHITAKAPGAPAAALCSQSQPAFLGTSSTSLTSSPHSGTSKR	540	
Db	481	SSVKSPSMSTLGHSTPRNLHITAKAPGAPAAALCSQSQPAFLGTSSTSLTSSPHSGTSKR	540	
Qy	541	RRVTDERELRIPLYGHQWRETRIRNFGRLQGEVAYYAPCCKKLQYQPEVILKYLNRNGIM	600	
Db	541	RRVTDERELRIPLYGHQWRETRIRNFGRLQGEVAYYAPCCKKLQYQPEVILKYLNRNGIM	600	
Qy	601	DISRDNFSFAKIRVGDIFYEARDGPQBWQCLKEEDVIPRAMEGRGRPPNPDORA	660	

601	Db	DISRDNFSFSAKURVGDGYFYBARQDQBMQWCLLKEEDVPIPRIAMEGRGRPPNPDRQRA	660
661	Qy	REESRMERRRGRPPNVGNAEFLDNADAKLLRKLQAOEIAFQAQAIKLLRKLQKOEQARVA	720
661	Db	REESRMERRRGRPPNVGNAEFLDNADAKLLRKLQAOEIAFQAQAIKLLRKLQKOEQARVA	720
721	Qy	KEAKKQQAIIWAAEBKRRKQKQIIKIMQOEKIKRIQQIRMEKELRAQOILBAKXKKKEEAA	780
721	Db	KEAKKQQAIIWAAEBKRRKQKQIIKIMQOEKIKRIQQIRMEKELRAQOILBAKXKKKEEAA	780
781	Qy	NAKLLAEKRIKKEKEMRROQAVLLKHOERERRRQHMLMKAMEARKKABEKRLKQEKRD	840
781	Db	NAKLLAEKRIKKEKEMRROQAVLLKHOERERRRQHMLMKAMEARKKABEKRLKQEKRD	840
841	Qy	EKRLINKERKLEORRLELEMAKELKPNEDMCLADQKPLPELPRIPGLVLSGSTFSDCLMW	900
841	Db	EKRLINKERKLEORRLELEMAKELKPNEDMCLADQKPLPELPRIPGLVLSGSTFSDCLMW	900
901	Qy	VQFLRNPFGKVLGFDVNI DVPNLSVLQEGLLNIGDSMGEVODLIVRLLSAAVCDPGLITGY	960
901	Db	VQFLRNPFGKVLGFDVNI DVPNLSVLQEGLLNIGDSMGEVODLIVRLLSAAVCDPGLITGY	960
961	Qy	KAKTALGEHLNLGVNRDNYSEILQIFMEAHCCQOTELTESLTKTKAQAHTPAOKASVLAF	1020
961	Db	KAKTALGEHLNLGVNRDNYSEILQIFMEAHCCQOTELTESLTKTKAQAHTPAOKASVLAF	1020
1021	Qy	LINELACSKVWSEIDKNIDYMSNLRRDKWVVEGKRLKLI IHAKTGKEDTSGGIDLGE	1080
1021	Db	LINELACSKVWSEIDKNIDYMSNLRRDKWVVEGKRLKLI IHAKTGKEDTSGGIDLGE	1080
1081	Qy	EQHPLGTPPTGRRKRRKGGSDYDDDDDDSDQGDDEDEDEKEDQOKKTKDICEDED	1140
1081	Db	EQHPLGTPPTGRRKRRKGGSDYDDDDDDSDQGDDEDEDEKEDQOKKTKDICEDED	1140
1141	Qy	EGDQASVLELEKQIEKLSKQSOQYRKLPDASHLSRVNFGPDRVRRRWIILPRCGGIF	1200
1141	Db	EGDQASVLELEKQIEKLSKQSOQYRKLPDASHLSRVNFGPDRVRRRWIILPRCGGIF	1200
1201	Qy	VEGMESGEGLTEAKREKLLKKAESVQIKEMEPETSGDSLNCSENTDHCQEKEDLKEKONT	1260
1201	Db	VEGMESGEGLTEAKREKLLKKAESVQIKEMEPETSGDSLNCSENTDHCQEKEDLKEKONT	1260
1261	Qy	NLFLOKPGSFKLSKLELVAKMPPSEVMPTPKPNAGANGCTLSYQNSGKHSLSGVSTAT	1320
1261	Db	NLFLOKPGSFKLSKLELVAKMPPSEVMPTPKPNAGANGCTLSYQNSGKHSLSGVSTAT	1320
1321	Qy	QSNVEKADSNLFWNTGSSGGPKGYSLPNDQLLKLTLEKQRQWFSLLPRTPCDDTSLTHA	1380
1321	Db	QSNVEKADSNLFWNTGSSGGPKGYSLPNDQLLKLTLEKQRQWFSLLPRTPCDDTSLTHA	1380
1381	Qy	DMSTASLVT PQSOPPSKSPSPTPAPLGSSAONPVGLNPPALSPLOVKGGVSMGLQFCGW	1440
1381	Db	DMSTASLVT PQSOPPSKSPSPTPAPLGSSAONPVGLNPPALSPLOVKGGVSMGLQFCGW	1440
1441	Qy	PTGVVTSNIPFTLSVPFSLGSLGSEGNNGNSFLTNSVASSKSPVPQNEKATSAQPAAV	1500
1441	Db	PTGVVTSNIPFTLSVPFSLGSLGSEGNNGNSFLTNSVASSKSPVPQNEKATSAQPAAV	1500
1501	Qy	EVAKPVDVFPSPKPIPEMQFGWRRIIDPEDLKALLKVLHURGIREKALQKQIOKHLDYIT	1560
1501	Db	EVAKPVDVFPSPKPIPEMQFGWRRIIDPEDLKALLKVLHURGIREKALQKQIOKHLDYIT	1560
1561	Qy	QACLKQNDVAIIELNNEENQVTRDIVENWVSVEEQAMEMDLSVLQVQVEDLERRVASASLQ	1620
1561	Db	QACLKQNDVAIIELNNEENQVTRDIVENWVSVEEQAMEMDLSVLQVQVEDLERRVASASLQ	1620
1621	Qy	VKGWMCPEPASEREDLVYFHKSFYTKCKEHDGEFTGEDESSAHLERKSDNPLDIAVTR	1680
1621	Db	VKGWMCPEPASEREDLVYFHKSFYTKCKEHDGEFTGEDESSAHLERKSDNPLDIAVTR	1680
1681	Qy	LADLERNIERRIEDTAPGLRVWRRLSEARSAAQVALCTIQLOKSIAMEKGSIMKYVCQI	1740
1681	Db	LADLERNIERRIEDTAPGLRVWRRLSEARSAAQVALCTIQLOKSIAMEKGSIMKYVCQI	1740

```
QY 1741 CRKGNRELLLLDCGCDKGCHTYCHRPKITTTIPGDWFCPACIAKASQTLIKIKLHVKG 1800
Db 1741 CRKGNRELLLLDCGCDKGCHTYCHRPKITTTIPGDWFCPACIAKASQTLIKIKLHVKG 1800
QY 1801 KKTNESKKGKVTTLTGDTDEDSASTSSSLKRGNKDLQKRWEENTSINLSKQESFTSVK 1860
Db 1801 KKTNESKKGKVTTLTGDTDEDSASTSSSLKRGNKDLQKRWEENTSINLSKQESFTSVK 1860
QY 1861 KPKRDDSXLALCSMILTEMETHEDAMPFLLPVNLKLVPGYKVKIKKPMDFSTIREKLSS 1920
Db 1861 KPKRDDSXLALCSMILTEMETHEDAMPFLLPVNLKLVPGYKVKIKKPMDFSTIREKLSS 1920
QY 1921 GOYPNLETFALDVLRFVDFNCETFEDEDDSDIGRAGHNMKRYPEKKWTDTFKVS 1972
Db 1921 GOYPNLETFALDVLRFVDFNCETFEDEDDSDIGRAGHNMKRYPEKKWTDTFKVS 1972

RESULT 3
US-10-702-148-21
; Sequence 21, Application US/10702148
; Publication No. US20040063145A1
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042002
; CURRENT APPLICATION NUMBER: US/10/702,148
; PRIOR FILING DATE: 2003-11-05
; PRIOR APPLICATION NUMBER: PRIOR APPLICATION NUMBER: US/09/839,479
; PRIOR FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCR/JF98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 21
; LENGTH: 1972
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-702-148-21

Query Match 100.0%; Score 10138; DB 15; Length 1972;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1972; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGOTKSTSSGGNRKCNQEQSNQPLDARVDKIKDKPRKAMSSSSNSDSGTSSTDS 60
Db 1 MGOTKSTSSGGNRKCNQEQSNQPLDARVDKIKDKPRKAMSSSSNSDSGTSSTDS 60
QY 61 SEG1SSSDSDLEEDDEEDSDIESRDDSDSEAOHKSNNVLLHGISDPKADGOKA 120
Db 61 SEG1SSSDSDLEEDDEEDSDIESRDDSDSEAOHKSNNVLLHGISDPKADGOKA 120
QY 121 TEKAQEKRIHQPLPLAFESQTHSFOSQOQOVLSQQLPFIQSSQAKEESVNKHTSVIQ 180
Db 121 TEKAQEKRIHQPLPLAFESQTHSFOSQOQOVLSQQLPFIQSSQAKEESVNKHTSVIQ 180
QY 181 STGLVSNVKPLSLVNOAKKFTYMKLIIVSPDVLKAGNKNTSEESSLLTSELRSREQYKQ 240
Db 181 STGLVSNVKPLSLVNOAKKFTYMKLIIVSPDVLKAGNKNTSEESSLLTSELRSREQYKQ 240
QY 241 APPSOLKKQESSKSLKVIKVAALSNPKATSSSPAHPKQTLNHNHNPFLTALLGNHOPNG 300
Db 241 APPSOLKKQESSKSLKVIKVAALSNPKATSSSPAHPKQTLNHNHNPFLTALLGNHOPNG 300
QY 301 VIQSVIQEAPLALTTKTKQSKINENIAAASSTPFSSPVNLTSGRRTPGNTQTPWPSAS 360
Db 301 VIQSVIQEAPLALTTKTKQSKINENIAAASSTPFSSPVNLTSGRRTPGNTQTPWPSAS 360
```

```
QY 361 PILHSQCKEKAVSNVNPVKTOHSHHPAKSLVEQFRGTDSIPSSKQSDSENEDEEEDDE 420
Db 361 PILHSQCKEKAVSNVNPVKTOHSHHPAKSLVEQFRGTDSIPSSKQSDSENEDEEEDDE 420
QY 421 BEDEDEDEDDSDSQSDSNSSESDTEGSEEEEDDDKQDQSDSDTEGEKTSMKLNKTT 480
Db 421 BEDEDEDEDDSDSQSDSNSSESDTEGSEEEEDDDKQDQSDSDTEGEKTSMKLNKTT 480
QY 481 SSVKSPMSLGTGHTPPNLHIAKAPGAPAAALCESQSOPALPGTSSSTLTSSPHSGTSCR 540
Db 481 SSVKSPMSLGTGHTPPNLHIAKAPGAPAAALCESQSOPALPGTSSSTLTSSPHSGTSCR 540
QY 541 RRVTDRELRIPILEYGWORETRIRNFGRLQGEVAYVAPCGKKLQYQPEVILKYLNRNIGM 600
Db 541 RRVTDRELRIPILEYGWORETRIRNFGRLQGEVAYVAPCGKKLQYQPEVILKYLNRNIGM 600
QY 601 DISRDNFSFAKIRVGDPEYARDGPQEMQWCLLKEEDVIPRIRAMEGRGRPPNPDQRA 660
Db 601 DISRDNFSFAKIRVGDPEYARDGPQEMQWCLLKEEDVIPRIRAMEGRGRPPNPDQRA 660
QY 661 REESRMRRRKGRPPNVGNAEFLDNADAKLRLKLAQAEIARQAAQIKLLRLKLOQEOARVA 720
Db 661 REESRMRRRKGRPPNVGNAEFLDNADAKLRLKLAQAEIARQAAQIKLLRLKLOQEOARVA 720
QY 721 KEAKKQQAIAWAAEBKQKQEQIKIMKQOEKIKRIQIIRWEKELRAQOILEAKKKKEEAA 780
Db 721 KEAKKQQAIAWAAEBKQKQEQIKIMKQOEKIKRIQIIRWEKELRAQOILEAKKKKEEAA 780
QY 781 NAKLLEAEKRIKKEEMRRQQAVALLKHOERERRRRRHHMLMKAMEARKKAEERLQKEKRD 840
Db 781 NAKLLEAEKRIKKEEMRRQQAVALLKHOERERRRRRHHMLMKAMEARKKAEERLQKEKRD 840
QY 841 EKRLNKRKLEBORLELEMAKELKKPNEDMCIAQOKPLPELPRIPLGLVLSGTSFSDCLMV 900
Db 841 EKRLNKRKLEBORLELEMAKELKKPNEDMCIAQOKPLPELPRIPLGLVLSGTSFSDCLMV 900
QY 901 VOFLRNFGKVLGPDVNDVNLVLOEGLNLNIGSMGEVQDOLLVRLLSAAVCDPLGITY 960
Db 901 VOFLRNFGKVLGPDVNDVNLVLOEGLNLNIGSMGEVQDOLLVRLLSAAVCDPLGITY 960
QY 961 KAKTALGEHLNLVGNVRDNLVSEILQIFMEAHCGQTELTESLTKTFAQHTPAQKASVLAF 1020
Db 961 KAKTALGEHLNLVGNVRDNLVSEILQIFMEAHCGQTELTESLTKTFAQHTPAQKASVLAF 1020
QY 1021 LINELACSKSVVSEIDKNIDYMSNLRRDKWVVEGKLRLRIIHAKKTGKRTSGSIDLGE 1080
Db 1021 LINELACSKSVVSEIDKNIDYMSNLRRDKWVVEGKLRLRIIHAKKTGKRTSGSIDLGE 1080
QY 1081 BOHPLGTPTGKRKRERRKGGSDYDDDDDDDDGDEDEDEKQKKTIDICRED 1140
Db 1081 BOHPLGTPTGKRKRERRKGGSDYDDDDDDDDGDEDEDEKQKKTIDICRED 1140
QY 1141 EGDQAAASVEELEKQIEKLSKQOQYRRKLFDASHLSLRVFMFGPDYRRRYWILPRCGGIF 1200
Db 1141 EGDQAAASVEELEKQIEKLSKQOQYRRKLFDASHLSLRVFMFGPDYRRRYWILPRCGGIF 1200
QY 1201 VEGMESGEGLEBEIAKEREKLLKASVOIKEEMFETSGDSLNCNSNTDHCEQKEDLKEKNT 1260
Db 1201 VEGMESGEGLEBEIAKEREKLLKASVOIKEEMFETSGDSLNCNSNTDHCEQKEDLKEKNT 1260
QY 1261 NLFLQKPGSFSLKLEVAQKMPPESEVMTPKPNAGANGCTLSYQNSGKSHISLGSVQSTAT 1320
Db 1261 NLFLQKPGSFSLKLEVAQKMPPESEVMTPKPNAGANGCTLSYQNSGKSHISLGSVQSTAT 1320
QY 1321 QSNVEKADSNLNFNTGSGPGKIFYSPLPNDQLLTKTEKNRQWFSLLPRTPCDDTSLTHA 1380
Db 1321 QSNVEKADSNLNFNTGSGPGKIFYSPLPNDQLLTKTEKNRQWFSLLPRTPCDDTSLTHA 1380
QY 1381 DMSTASLVTPOSPQPSKSPSPTAPLAGSSAQNVLGNPFALSPLOVKGVSMMGLQFCGW 1440
Db 1381 DMSTASLVTPOSPQPSKSPSPTAPLAGSSAQNVLGNPFALSPLOVKGVSMMGLQFCGW 1440
```

QY 1441 PTGVVTSNIPFTLSVPSLGSGLSEGNCSFLTNSVASSKSPVPQNEKATSAQPAAV 1500
Db 1441 PTGVVTSNIPFTLSVPSLGSGLSEGNCSFLTNSVASSKSPVPQNEKATSAQPAAV 1500
QY 1501 EVAKPVDPPSPKPIPEEMQFGWRIIDPEDLKALLKVLHLRGIREKALQKQIKHLDYIT 1560
Db 1501 EVAKPVDPPSPKPIPEEMQFGWRIIDPEDLKALLKVLHLRGIREKALQKQIKHLDYIT 1560
QY 1561 QACLKXKDVAIITELNENEENQVTRDIVENWSVEEQAMEMDLVLOQVEDLERRVASASIQ 1620
Db 1561 QACLKXKDVAIITELNENEENQVTRDIVENWSVEEQAMEMDLVLOQVEDLERRVASASIQ 1620
QY 1621 VKGWCPEPASREDLVFPEHKSFTKLCHEHGBTEGDESSAHALERKSDNPLDIATVR 1680
Db 1621 VKGWCPEPASREDLVFPEHKSFTKLCHEHGBTEGDESSAHALERKSDNPLDIATVR 1680
QY 1681 LADLERNERREREDTAPGLRVRRAALSEARSAOVALCIIQLOKSIAWEKSIMKVYCOI 1740
Db 1681 LADLERNERREREDTAPGLRVRRAALSEARSAOVALCIIQLOKSIAWEKSIMKVYCOI 1740
QY 1741 CRKGDNEELLCDGCDKGCHTYCHRPKITTTIPDGDWFCPACIAKASGOTLKI KKLHVKG 1800
Db 1741 CRKGDNEELLCDGCDKGCHTYCHRPKITTTIPDGDWFCPACIAKASGOTLKI KKLHVKG 1800
QY 1801 KKTNESKKGKVTLLTGDTEDEDSASTSSSLKRGKNDLQKRMEENTSINLSKQESFTSVK 1860
Db 1801 KKTNESKKGKVTLLTGDTEDEDSASTSSSLKRGKNDLQKRMEENTSINLSKQESFTSVK 1860
QY 1861 KPRKDDSDALCMLTETHEADWPELLVNLKLVPGYKVKIKKPMDFSTIREKLSS 1920
Db 1861 KPRKDDSDALCMLTETHEADWPELLVNLKLVPGYKVKIKKPMDFSTIREKLSS 1920
QY 1921 GOYPNLETALDVLVDFNCETFNEDSDSIIGRAGHNMRYFEKKWTDTFKVS 1972
Db 1921 GOYPNLETALDVLVDFNCETFNEDSDSIIGRAGHNMRYFEKKWTDTFKVS 1972

RESULT 4

US-09-839-479-71
; Sequence 71, Application US/09839479
; Publication No. US20020039779A1
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042092
; CURRENT APPLICATION NUMBER: US/09/839,479
; CURRENT FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 71
; LENGTH: 1969
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-839-479-71

Query Match 98.9%; Score 10022.5; DB 9; Length 1969;
Best Local Similarity 99.4%; Pred. No. 0;
Matches 1962; Conservative 2; Mismatches 4; Indels 5; Gaps 5;
QY 1 MGQTKSTSSGGNRKCNQEQSKNQPLDARVDKIKDKKPRKKAMESSNSDSGTSSTDS 60
Db 1 MGQTKSTSSGGNRKCNQEQSKNQPLDARVDKIKDKKPRKKAMESSNSDSGTSSTDS 60
QY 61 SEGISSSDDLLEDEEEDQSI ESEDDSDSESAQHKSNNQVLLHGISDPKADGQKA 120

Db 61 SEGISSSDDLLEDEEEDQSI ESEDDSDSESAQHKSNNQVLLHGISDPKADGQKA 120
QY 121 TEKAQEKRIHQPLPLAFESQTHSFOSQOQPOVLSQLPFI FQSSQAQKEESVNGKHTSVIQ 180
Db 121 TEKAQEKRIHQPLPLAFESQTHSFOSQOQPOVLSQLPFI FQSSQAQKEESVNGKHTSVIQ 180
QY 181 STGLVSNVKSPLSLVNOAKKETVMKLI VPSPDVLKAGNKNTSESSLLTSELSESKRQYKQ 240
Db 181 STGLVSNVKSPLSLVNOAKKETVMKLI VPSPDVLKAGNKNTSESSLLTSELSESKRQYKQ 240
QY 241 APPSOLKQESKSLKKVIAALSNPKATSSSPAHPKQTL ENNHNPFPFLTNALLGNHQPNG 300
Db 241 APPSOLKQESKSLKKVIAALSNPKATSSSPAHPKQTL ENNHNPFPFLTNALLGNHQPNG 300
QY 301 VIQSVIQEAPLALTTTKMQSKINENIAAASSTPFSSPVNLSTSGRRTPGNTPVMPAS 360
Db 301 VIQSVIQEAPLALTTTKMQSKINENIAAASSTPFSSPVNLSTSGRRTPGNTPVMPAS 360
QY 361 PILHSQGEKAVSNVNPVKTOHSHHPAKSLVEQPRGTOSDI PSSKSDSDSNEDEEEDDE 420
Db 361 PILHSQGEKAVSNVNPVKTOHSHHPAKSLVEQPRGTOSDI PSSKSDSDSNEDEEEDDE 420
QY 421 EDEEDDDDDSDSDSSESDTSGSEBEEEDDDKQDSDSDTGEKTSMLKNTT 480
Db 421 EDEEDDDDDSDSDSSESDTSGSEBEEEDDDKQDSDSDTGEKTSMLKNTT 480
QY 481 SSVKSPMSLSTGHSPTPNLHIAKAPGSAPALCSESQSPAF LGTSSSTLTSSPHSGTSKR 540
Db 480 SS-KSPMSLSTGHSPTPNLHIAKAPGSAPALCSESQSPAF LGTSSSTLTSSPHSGTSKR 538
QY 541 RRVTDERBLRIPLEYGWQRETRIRNFGGRLOGEVAYAPCGKKLQYPEVI KYLSRNGIM 600
Db 539 RRVTDERBLRIPLEYGWQRETRIRNFGGRLOGEVAYAPCGKKLQYPEVI KYLSRNGIM 598
QY 601 DISRDNFSFAKIRVGYDEYARDGQEMQWCLLKEEDVI PRIRAMEGRGRPNPDORQA 660
Db 599 DISRDNFSFAKIRVGYDEYARDGQEMQWCLLKEEDVI PRIRAMEGRGRPNPDORQA 658
QY 661 REESMRRRKGRPPNVGNAEFLDNADAKLLRLKLAQEIARQAQIKLRLKLOQEQARVA 720
Db 659 REESMRRRKGRPPNVGNAEFLDNADAKLLRLKLAQEIARQAQIKLRLKLOQEQARVA 717
QY 721 KEAKKQQAIAAAEKKRQKEQIKIMKQKEKIKRIQOIRMEKELRAQOILEAKKKKKEBAA 780
Db 718 KEAKKQQAIAAAEKKRQKEQIKIMKQKEKIKRIQOIRMEKELRAQOILEAKKKKKEBAA 777
QY 781 NAKLLEAEKRIKEKEMRRQQAIVLLKHQERERRRQHMLMKAMEARKKAEKERLQKED 840
Db 778 NAKLLEAEKRIKEKEMRRQQAIVLLKHQERERRRQHMLMKAMEARKKAEKERLQKED 837
QY 841 EKRLNKERKL BORRLEMAKELKPNEDMCLADOKPLPELPRI PGLVLSGTSFSDCLMV 900
Db 838 EKRLNKERKL BORRLEMAKELKPNEDMCLADOKPLPELPRI PGLVLSGTSFSDCLMV 897
QY 901 VQFLNFGKVLGFDVNI DVPNLSVLQEG-LLNIGDSMGEVQDQLVRLLSAAVCDPGLITG 959
Db 898 VQFLNFGKVLGFDVNI DVPNLSVLQEG-LLNIGDSMGEVQDQLVRLLSAAVCDPGLITG 957
QY 960 YKAKTALGEHLNVLGVNRDQVSEILQIFMEAHCGQTEL TESLKTVAFOAHTPAQKASVLA 1019
Db 958 YKAKTALGEHLNVLGVNRDQVSEILQIFMEAHCGQTEL TESLKTVAFOAHTPAQKASVLA 1016
QY 1020 FLINELACKSVVSEIDKNIDVMSNLRDKWVVEGKRLKRIIHAHKTGKRTSGGIDIG 1079
Db 1017 FLINELACKSVVSEIDKNIDVMSNLRDKWVVEGKRLKRIIHAHKTGKRTSGGIDIG 1076
QY 1080 BEQHPGLTPTGKRKRRKGGSDYDDDDDDSDSDGDEDEDEDEDEDEDEDEDEDEDEDE 1139
Db 1077 EEQHPGLTPTGKRKRRKGGSDYDDDDDDSDSDGDEDEDEDEDEDEDEDEDEDEDEDE 1136
QY 1140 DEGQAAASVEELEKQIEKLSKQOQSYRRKLFDAHSLRSVMFGPDYRRRYWILPRCGGI 1199
Db 1137 DEGQAAASVEELEKQIEKLSKQOQSYRRKLFDAHSLRSVMFGPDYRRRYWILPRCGGI 1196

QY 1200 FVEGMESGEGLEEIAKREKLUKKAESVOI KEEMFETSGDSLNCNNTDCEBOKEDLKEKDN 1259
DB 1197 FVEGMESGEGLEEIAKREKLUKKAESVOI KEEMFETSGDSLNCNNTDCEBOKEDLKEKDN 1256
QY 1260 TNLFLQKPGSFKLSKLLEVAKMPPESEVMTPKPNAGANGCTLSYQNSGKHSLSGVQSTA 1319
DB 1257 TNLFLQKPGSFKLSKLLEVAKMPPESEVMTPKPNAGANGCTLSYQNSGKHSLSGVQSTA 1316
QY 1320 TQSNVEKADSNLNFNTGSGGCKFYSLPNDQLLKTLEKRNQWPSLLPRTPCDDTSLTH 1379
DB 1317 TQSNVEKADSNLNFNTGSGGCKFYSLPNDQLLKTLEKRNQWPSLLPRTPCDDTSLTH 1376
QY 1380 ADMSTASLVTPQSOPPSKPSPTAPLGGSSAQNPVGLNPFALSPLOVKGVGVMGLOFCG 1439
DB 1377 ADMSTASLVTPQSOPPSKPSPTAPLGGSSAQNPVGLNPFALSPLOVKGVGVMGLOFCG 1436
QY 1440 WPTGVVTSNIPTFTLSVPISGLGSLSEGNNGNSFLTNSVASSKSBSPPVQNEKATSAQPA 1499
DB 1437 WPTGVVTSNIPTFTLSVPISGLGSLSEGNNGNSFLTNSVASSKSBSPPVQNEKATSAQPA 1496
QY 1500 VEVAKPVDPSPKPIPEBMQGWRIIDPDLKALLKVLHLRGIREKALQIOKHLDIYI 1559
DB 1497 VEVAKPVDPSPKPIPEBMQGWRIIDPDLKALLKVLHLRGIREKALQIOKHLDIYI 1556
QY 1560 TOACLNKNDVAIIELENENEQVTRDIVENNSVEEQAMEMDLVLOQVEDLERRVASASL 1619
DB 1557 TOACLNKNDVAIIELENENEQVTRDIVENNSVEEQAMEMDLVLOQVEDLERRVASASL 1616
QY 1620 QVKGWMCPEPASEREDLVYFEHKSFTKLCKEHDEGFTGEDESSAHALERKSDNPLDIATV 1679
DB 1617 QVKGWMCPEPASEREDLVYFEHKSFTKLCKEHDEGFTGEDESSAHALERKSDNPLDIATV 1676
QY 1680 RLADLERNIERIBEDIAPGLRVWRRLSEARSAAQVALCTIOQOKSIWAEKSIWKVYQC 1739
DB 1677 RLADLERNIERIBEDIAPGLRVWRRLSEARSAAQVALCTIOQOKSIWAEKSIWKVYQC 1736
QY 1740 ICRGDNELLLDCGDKGCHTYCHRPKITTIPDGMWFCPACIAKASGGTLKKLHVK 1799
DB 1737 ICRGDNELLLDCGDKGCHTYCHRPKITTIPDGMWFCPACIAKASGGTLKKLHVK 1796
QY 1800 GKKTNESKKGKVTITGTDTEDEDSASTSSSLKRGKNOIQKRWBEENTSINLSKQESFTSV 1859
DB 1797 GKKTNESKKGKVTITGTDTEDEDSASTSSSLKRGKNOIQKRWBEENTSINLSKQESFTSV 1856
QY 1860 KKPKRDDSKDALCSMIITEMETHEDAWPFLPVNLKLVPGYKVKIKPMDFTSTIREKLS 1919
DB 1857 KKPKRDDSKDALCSMIITEMETHEDAWPFLPVNLKLVPGYKVKIKPMDFTSTIREKLS 1916
QY 1920 SGQYPNLETFALDVLVFDNCETNEDESDIGRAGHNNRKYFEKKWTDTEKVS 1972
DB 1917 SGQYPNLETFALDVLVFDNCETNEDESDIGRAGHNNRKYFEKKWTDTEKVS 1969

RESULT 5

us-10-376-537-72
; Sequence 72, Application US/10376537
; Publication No. US20030224405A1
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; CURRENT APPLICATION NUMBER: US/10/376,537
; FILE REFERENCE: 06501-042001
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US/09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 73

; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 72
; LENGTH: 1969
; TYPE: PRT
; ORGANISM: Homo sapiens
us-10-376-537-72

Query Match 98.9%; Score 10022.5; DB 15; Length 1969;
Best Local Similarity 99.4%; Pred. No. 0;
Matches 1962; Conservative 2; Mismatches 4; Indels 5; Gaps 5;

QY 1 MGQTKTSSSGGNKCKNQESKQNPDLARVDKIKDKPRKKAMESSNSDSGTSPTS 60
DB 1 MGQTKTSSSGGNKCKNQESKQNPDLARVDKIKDKPRKKAMESSNSDSGTSPTS 60
QY 61 SEG1SSSDSDLEDEDEEDSIESEEDSDSSESAQHKSNNQVLLHG1SDPKADGOKA 120
DB 61 SEG1SSSDSDLEDEDEEDSIESEEDSDSSESAQHKSNNQVLLHG1SDPKADGOKA 120
QY 121 TEKAQEKRIHOPLAPAFESQTHSFQSQOKQVLSQQLPFIQSSQAKEESVKNKHTSVIQ 180
DB 121 TEKAQEKRIHOPLAPAFESQTHSFQSQOKQVLSQQLPFIQSSQAKEESVKNKHTSVIQ 180
QY 181 STGLVSNVKPLSLVNQAKKETMYMLIVPSPDVLKAGNKNTSESSLLTSELRSKREQYKQ 240
DB 181 STGLVSNVKPLSLVNQAKKETMYMLIVPSPDVLKAGNKNTSESSLLTSELRSKREQYKQ 240
QY 241 APPSOLKKQESSKSLKVVIAALSNPKATSSSPAHPKQTLNHNHPNPLTNALLGNHQPNG 300
DB 241 APPSOLKKQESSKSLKVVIAALSNPKATSSSPAHPKQTLNHNHPNPLTNALLGNHQPNG 300
QY 301 VIQSVIOEAPLALTTKTMQSKINENIAAASSTPSSPVLNSTSGRRTPGQTPWPSAS 360
DB 301 VIQSVIOEAPLALTTKTMQSKINENIAAASSTPSSPVLNSTSGRRTPGQTPWPSAS 360
QY 361 PILHSQKKEKAVSNVNPVKTQHHSHPAKSLVEQFRGTDSDIPSSKQSDSENEDEEDE 420
DB 361 PILHSQKKEKAVSNVNPVKTQHHSHPAKSLVEQFRGTDSDIPSSKQSDSENEDEEDE 420
QY 421 EDEDEDEDDSDSQSESNSSESDTEGSEEDDDDKQDESDDTEGETSKMLNKTT 480
DB 421 EDEDE-DEDDSDSQSESNSSESDTEGSEEDDDDKQDESDDTEGETSKMLNKTT 479
QY 481 SSVKSPMSLTGHSHTPNLHIAPGAPAPALCSSESQSPAFLGTSSTLTSSPHSGTSKR 540
DB 480 SS-KSPMSLTGHSHTPNLHIAPGAPAPALCSSESQSPAFLGTSSTLTSSPHSGTSKR 538
QY 541 RRVTDERELRIPLEYGMQRETRIRNFGRLQGEVAYYAPCGKLRQYPEVIKYLNRNGIM 600
DB 539 RRVTDERELRIPLEYGMQRETRIRNFGRLQGEVAYYAPCGKLRQYPEVIKYLNRNGIM 598
QY 601 DISRDNFSFAKIRVGDIFYEARDGPQEMQWCLLKEEDVIPRAMEGRGRPPNDRORA 660
DB 599 DISRDNFSFAKIRVGDIFYEARDGPQEMQWCLLKEEDVIPRAMEGRGRPPNDRORA 658
QY 661 REESMRRRKGRPPNVGNABFLDNADAKLRLKAOETARQAOIKLLRLKQKQBOQARVA 720
DB 659 REESMRRRKGRPPNVGNABFLDNADAKLRLKAOETARQAOIKLLRLKQKQBOQARVA 717
QY 721 KEAKKQQAIAAAEEKQKEQIKIMKQOEKIKRIQIIRMEKELRAQQILEAKKKKEBAA 780
DB 718 KEAKKQQAIAAAEEKQKEQIKIMKQOEKIKRIQIIRMEKELRAQQILEAKKKKEBAA 777
QY 781 NAKLLEAKRIKEKEMRRQQAIVLLKHORERRRRRHHMLMKAMEARKKAAEERLQKQKRD 840
DB 778 NAKLLEAKRIKEKEMRRQQAIVLLKHORERRRRRHHMLMKAMEARKKAAEERLQKQKRD 837
QY 841 EKRLNKRKLQRLLEMAKELKPNEDMCLADQKPLPELPRIPGLVLSGSTFSDCLMV 900
DB 838 EKRLNKRKLQRLLEMAKELKPNEDMCLADQKPLPELPRIPGLVLSGSTFSDCLMV 897
QY 901 VQFLRNFQKVLGFDVNDVNPVLSVLOEG-LLNIGDSMGEVODLLVRLLSAAVCPGLITG 959
|||||

RESULT 9

US-10-376-537-71
; Sequence 71, Application US/10376537
; Publication No. US20030224405A1
; GENERAL INFORMATION:

; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042001
; CURRENT APPLICATION NUMBER: US/10/376,537
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US/09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 71
; LENGTH: 1876
; TYPE: PRT
; ORGANISM: Homo sapiens

US-10-376-537-71

Query Match 24.6%; Score 2493.5; DB 15; Length 1876;

Best Local Similarity 32.1%; Pred. No. 2.1e-112;

Matches 675; Conservative 343; Mismatches 674; Indels 414; Gaps 66;

QY 43 MESSNSDSD-----SOTS-SDTSSGSISSSDSDDEEEDDEEEDSIESED 88
DB 1 MEMEANEAHFNFTGLPPAPAAAGLKPSPSGEGLYTNGS---PHNFPQGGKSL--NGD 55
QY 89 DSDSDSEBAQKSNNOVLHGISDPKADQKATEKAQEKRIHOPPLAFE-----SOTHS 143
DB 56 VNVNGLSTVSHTTSGIL-----NSAPHSSSTSHLHP--SVAYDCLMNSQYPS 103
QY 144 FQ--SQKQPOVLQO-----QLPFIQSSQAKESVNKHTSVIQ-----STGLVSNVKP 190
DB 104 ANPGSNLKDPLLSQFGQYPLNGILGSGRQSPSPSHNTNLRAGSQKFWANGTHS---P 160
QY 191 LSLVNAQAKETMYKL-----IVPSDDVLKAGNKNTSESSLLTSELRSKREQ-- 237
DB 161 MGL-NFDSQELYDSDPDQNFEEVCSGIHPDEAAEKEMTSVVAENGTGLVCSLELEEEQPE 219
QY 238 -----YKQAFPSQLKQSSKLVIAALSNPKATSSSPAHPKOTLENNHPPLTNAL 292
DB 220 LKMCYNGSVP-----SVESLHQEVSLV--PDPTVSLDDP----- 254
QY 293 LGNHOPNGVIOQVIOEAPLALTTKMQSKINENIAAASSTPSSPVNLSTSGRRTPGNQ 352
DB 255 --SHLPD-----QLEDITI-----LSEDSLEFPNLSLAPEPVSGGLYGIDD 292
QY 353 TPVMPSPAILHSGQKEKAVSNVNPVK-----QHGHSPAKSLVQEPFGRTSDIPSSKD 407
DB 293 TELM-----GAEDKLPLEDSPVISALDCPSLNNATAFSLADSDSTSTISFASPT 342
QY 408 SEDSNDEDEEDDEEDDEDDSDSQSDSDNSSEDSTGSEEDDDDKQDESDDT 467
DB 343 SPPVLGSEVLQDNSFDLNGSDAEQEMETQSSDPPPSLTQPA-----PDQSSTIQ 393
QY 468 EGEKTSMLKNKTT-----SVKSPSMS----- 489
DB 394 LHPATSPAVPTTSPAVSLVSPASPEISPEVCPAASTVVSFVSPASSAVLPVAV 453
QY 490 -----LTHGHTSPNLHIAKAPCSA--PAALCESQSPAFLTGSSST-----LTSPPHSG 536
DB 454 SLEVPLTASVTSPPASVTPSAPAAFPPTASPAKDVSSFLETTADVEEITGEGLTAS--GSG 512
QY 537 TSKRRVTDRELRIPLEYGHQRETRINFGRLOGEVAYVAPCKKLRQYPEVIKYLRS 596
DB 513 DVMRRRIATPEBVLPLQHGWRVREIRIKGSHRWQGETWYIYGCPCCKRMKQPFPEVIKYLRS 572


```
Db 1104 AGIFVEGTEGTVLVEVIVKETSLSKVAHAASLNPALFSMKMELAGSNTTASSPARASR 1163
Qy 1252 EDLKEKDNLTNLQKPGFSKLSKLE---VAKMPPSEVETNP-KPNAGANGCTLSY-- 1304
Db 1164 -PLTKPG--PMQ--PREFKSPVRQDSEQPOAQLOPEAQLHVPAQPOQLOLOSHKG 1218
Qy 1305 ---QNSGKHSGSVQSTQSH---VEKADSNLNFN---TGSSGPKFYSPPLND--- 1350
Db 1219 FLEQSGSLSLQSQSHLSQSAFLSWLSQTOHSSLLSSSVLTPTDSSPGKL-DPAPSQPPE 1277
Qy 1351 ---QLLXTLTKKNQWESLLPRTCDTSLTHADWSTASLVTQSPQSPSPSTPAPL 1406
Db 1278 EPEDEASSPLQAFWNISQMPCN-----AAPTPLAVSEDOPTSPPOOL 1325
Qy 1407 GSS---AQNVPGLNPPALPQVKGGSMMGL---QFCGWFTGVVTSNIPFTLSVPVSLGSG 1461
Db 1326 ASSKPMNPSAANP--CSFQVP-SSTPLAGLAPKRRAGDPG-----EMPOSTPG 1371
Qy 1462 LGLSEGNNGNSFLTNSVASSKSSPVPQNEKATSAQAAVEVAKVPDVPSPKPIPEMQFG 1521
Db 1372 LQPKRRRK-----PESKF---FKQMEQRVLTQLTA-----QVPPPEMCSG 1409
Qy 1522 WRRITDPEDLKALLVHLRGIREKALQKQKHLDYITQACLKXKVAIIELENEENQ 1581
Db 1410 WMIIDPEMLDAMLKALHPRGIREKALHKLNRHDFLOEVCLR---PSADPIPEPRQLP 1466
Qy 1582 VTRDIVENWSVEQAMENDSLVLOQVEDLERRVASASLVQKGMWCEPASEREDLAVPEH 1641
Db 1467 AFQEGIMWSPEKTYETDLAVLQWVELEQVRVMSDQIRGWTCPSPDSTREDLACEH 1526
Qy 1642 KSFTKLCKEHGEFTGEDESSAHLERKSDNPLDIATVRLADLERNIERR----- 1691
Db 1527 LSDS---QEDITWRGPGREGI-APQRKTTNPLDLAVMLAALAEQNVKRRYLRELPWPTH 1581
Qy 1692 -----IEDIAPGLVWRALSEARSAQAVALCLOOLQKSTAW 1729
Db 1582 EWLKALLSTPNGAPEGTTTSEIYEITPRIMRQTLQRCSAAHVCLCLGHLERSIAW 1641
Qy 1730 EKSIMKVCQICRKGDNELLLCGDKGCHYCHRPKITTPDGMWFCPACIAKASGQ 1789
Db 1642 EKSVMKVTCLVCRKGDNEDEFLLLCGDCRGCHYCHRPKMEAVPGDNFCTVCLAQ---- 1697
Qy 1730 TLKIKLHVKGKKTWE---SKGKGVTLTGDTEDDSASTSSSLKRGKNDLQKRMEBNT 1846
Db 1698 -----QVEGFTQKPGPKGQK-----RKSGVSLNFSEGDCRRRVLKGRESAA 1744
Qy 1847 SINLSKQESFTSVKPP--KRDSDKDALCSMLITTEMETHEDAWPPLLPLVNLKVPYKVV 1904
Db 1745 GPRYSEERLSKRRPLSMRNHSHLTFCEIILMEMESHDAWPPLEFVNPRLVSGYRRI 1804
Qy 1905 IKKPMDFSTIREKLSGGVYNLETFALDVRLVFDNCETFNEDSDSDIGRAGHNMRYFEKK 1964
Db 1805 IKNPMDFSTMRERLURGGTSEEPFAADALLVDFNCQTFNEDDSEVGKAGHIMRRFFESR 1864
Qy 1965 WTDTPK 1970
Db 1865 WEEFYQ 1870
```

RESULT 11

```
US-09-839-479-13
; Sequence 13, Application US/09839479
; Publication No. US20020039779A1
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042002
; CURRENT APPLICATION NUMBER: US/09/839,479
; CURRENT FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
```

```
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 1878
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(1878)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-09-839-479-13
```

Query Match 24.5%; Score 2482.5; DB 9; Length 1878;
Best Local Similarity 32.2%; Pred. No. 7.3e-112;
Matches 680; Conservative 329; Mismatches 678; Indels 426; Gaps 68;

```
Qy 43 MESSNSDS-----SGTSSDTSSEGISSSDSDLEDEDEEEDQSISEESDDSDSESAQ 98
Db 1 MEMEANENDHFNFTGLPPAPASGLKSPS-----SCEGLYTNCSPNPFPQ 48
Qy 99 HKS-NNQVLLHG---ISDPKADG--QKATEKAQEKRIHQPLPAFE-----SQTHSFQ-- 145
Db 49 GKSLGDNVNVNGLSVSHSTTTSGILNSAPHSSSTSHLHP-SVAYDCLWNYSQYPSANPG 107
Qy 146 SQQKQPOVLSQQLPFIPOSSQAKESVKNKTSVSTQGLSVNVKPLSLVNQAKKXTYMKL 205
Db 108 SNLKDPPLLSQ-----FSGQYPLNGI-----LGGSRQPS----- 138
Qy 206 IVPSPDV-LKAGNKNTSESSLLTSELRSKREQYKAPPSOLKKQESSKSLKVAALSN 264
Db 139 --PHNTNLRAGSKFWANGTHSPMGLNFDQELYDSFPDQ-----NFEVCSGI-- 186
Qy 265 PKATSSPAHPKQTLNHNHPFLTNAL-----LGNHQP-----NG-----VIOSVTQ 307
Db 187 -----HPDEAAEKEMTSVVAENGTLVCSLEEXQPELKMCGYNGSVPSVESLHQ 237
Qy 308 EAPL-----ALTTKTKMSKINEN--IAAASSTPSSPNVLTSGRRTPGNTPMP 357
Db 238 EVSVLPDPTVSCLDLDPHLDPQLEDTPILSEDSLEPFNSLAPEVPVSGLYGIDDTLM- 296
Qy 358 SASPLHSQKEXKAVSNVNPVKT-----QHSHHPAKSLVEQFGTGDIDIPSSKDSN 412
Db 297 -----GAEKLPXDSPVISALDCPSLNNATAFSLADDSTSTISIFASPTSPVL 347
Qy 413 EDEEDDEEEDDEDDSDSQSSEDSNSESDETEGSEEDDDDDQDQDESDDTEGKT 472
Db 348 GESVLQDNSFDLNNGSDAEQEMETQSSDFPPLTQPA-----PDQSSTIQLHPAT 398
Qy 473 SMKLNKTTT-----SVKSPMS----- 489
Db 399 SPAVSTTSPAVSLVSPASPEISPEYCPAASTVSPAVFSVWSPASSAVLPAYSLVP 458
Qy 490 LTGHSTPRLNLI AKAPGSA-PAALCSQSQAFLGTSSST-----LTSSPHSGTSKRR 541
Db 459 LTASVTPSPKSPVTSAAAFTASPAKNDVSGFLETTADVEEITGEGITAS--GSDVMRR 517
Qy 542 RVTDERELRIPLEXGWQRETRIRNFGGRLOGEVAYYAPCGKKLQRYPEVIKYLNRNGIMD 601
Db 518 RIATPEEVRLPLQHGWRREVRKKGSHRWQGETWYYPGCGKMKQFPPEVIKYLNRNVHS 577
Qy 602 ISRNFSPSAKIRVGDIFYEARDGQEMQWCLLKEEDVIPRIAMEGRGRPPNPQRAR 661
Db 578 VRREHFSFSPMPVGDFFPEERDTPEGLQWVLSAEEIPSRIQAITGKRGRPRENTEKATK 637
Qy 662 EESRMRERKGRPPNVGNAEFLDNADAKLLRKLQAEIARQAQIKLLRKLQEQEARVAK 721
Db 638 EVPKVKRGRGRPPPKVKITELLNKNTDNRPLKLEAQE-----TLNEEDKAKIAK 685
```



```

Db      1700 -----QVEGFTQKPGPKRGQK-----RKSGYSLNFSGDDGRRRLVLK 1739
Qy      1843 -BENTSINLSKOEFTSVKPK-----RDSKDALALCSMILTEMETHEDAWPFLPLVLK 1897
Db      1740 GRESAAGPRYSEERLSPSKRRLSRNHSDLTTCIILMESHDAAPFXEPVNPRL 1799
Qy      1898 VPGYKVIKPNDFSTIREKLSGQYPNLETALDRLVDFNCETFNEDDSDIGRAGHNM 1957
Db      1800 VSGYRIIKPNDFSTIREKLSGQYPNLETALDRLVDFNCETFNEDDSDIGRAGHNM 1859
Qy      1958 RYFEKKWTDTEK 1970
Db      1860 RFFESRWEFFQ 1872

RESULT 14
US-10-087-192-663
; Sequence 663, Application US/10087192
; Publication No. US20020182586A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; APPLICANT: Engelhard, Eric K.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: CANCER
; FILE REFERENCE: 529452000122
; CURRENT APPLICATION NUMBER: US/10/087,192
; CURRENT FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 2059
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 663
; LENGTH: 1586
; TYPE: PR1
; ORGANISM: Mus musculus
US-10-087-192-663

```

```

Query Match      22.7%; Score 2300; DB 13; Length 1586;
Best Local Similarity 30.7%; Pred No. 4,3e-103;
Matches 599; Conservative 265; Mismatches 501; Indels 584; Gaps 52;

Qy      183 GLVSNVXPLSLVNQAK---ETMKLIVPSDVLKAGNKNTESESSLTSELRSKREQYK 239
Db      55 GLYNGSPNFPQOGKSLNGDVNGLSTVSHTTTSGILNSAPHSS-STSHLHPNVAYD 113
Qy      240 -----QAFPSQLKQESSKSLKVIATLALSNPKATSSPAHPKQTLNNHPNPLTNALLG 294
Db      114 CLWNYSQYPS-----ANPGNNLKDPLLSQFPFGQYP---LNGILGG 152
Qy      295 NHQPNQVISOVIOEAPLALTTKMQSKINENIAAAS-----STPFSSPVNLSTSGRRT-- 348
Db      153 NRQPS-----SPHNTNLRAGSEFWANGTQSPWGLNFDSELYD 192
Qy      349 ---PGNQTPVMP-----ASPIHSQKKEKAVSNVNPVKTHQHSHPAKSLVEQF- 395
Db      193 SFPDQNFVMPNGPPSPFTSPQTSPLMGSSIQTFAPSQDVS-----SDIHPDEAAKELT 247
Qy      396 -----RGT-----DSDIPSK-----DSDGSNDE-----BEDDEE 421
Db      248 SVVAENGTLVGLSLEBEQPELKMCGYNGSVSSVLSHQESVLVDPDVTVSCLDPSHL 307
Qy      422 EDEEDEDDESQSDSNSESQ-TEGSEBEDDDKQDESDDTGEKTSMKLNKTT 480
Db      308 PDQLEDTPILSEDSLEPFDLSAAAPVSGSLYGDIDDAELMGABD-----KLPLEGNPVI 361
Qy      481 SSVKSPSM-----SLTGHSTPRNLHIAKAPGSAPA-----ALCSSESQSPA 520
Db      362 SALDCPALSNANAFSLLADDQSASIFVSPISPVILGESVLQGSPEPSPAAFQTVSPA 421
Qy      521 FLGTSSSTLTSPH-----SGTSKRRRVTDRELRIPLLEYQWQRETRINFGG 568

```

```

Db      422 RKNVSSAPKARADRETTGGAVAVSGSDVLKRRJATPEEVRLPLQHGWRREVR1KKGSH 481
Qy      569 RLQGEVAYVAPCGKKLRQYPEVIKYLNRNGIMDISRDNFSFSAKIRVGDGYEARDQPSBM 628
Db      482 RWQGETWYGGCGKMKQFPPEVIKYLNRNVHSVPRHEHFSFSPMPVGDFFERDTPGL 541
Qy      629 QWCLLKEDVTIPRAMEGRGRPPNPRORAREESRMRKGRPPNVGNABFLNADAK 688
Db      542 QWVQLSABEIPSRIOAITGKGRPRNNEKAKNKEVPKVKRGRGRPPKIMPELLNKTNR 601
Qy      689 LLRLQAOEIAARQAQIKLLRKLOEQARVAKAEKQQAIAAAEKKQKQEQIKIMKQ 748
Db      602 LPKKLETOEI-----LSEDDKAKMTKKNKGM-----RQVQ----- 632
Qy      749 EKIKRIOIRMEKELRAQOILEAKKKKEEAAANAKLEAEKRIKEK---EMRRQQAALLKH 806
Db      633 -----RGESQTPVG--QARNKRQDIKSLKQKDTKKLKKLEBEQOQQAAIL--- 678
Qy      807 QERERRRHHMLMKAMEARKKAEKERLQKEKRDEKRLNKERKLEQRLELEMAKELKPP 866
Db      679 -----BEMKKP 684
Qy      867 NEDMCLADQKPLPELPRIPLGLVLSGSTFSDCLMVVOFLRNFCKVLGFVNDVPLSLVQ 926
Db      685 TEDMCLSDHQPLPDFTRIPLGLTSLSRASFSDCLTIVEFLHSFGKVLGFLDKDVPVSLVQ 744
Qy      927 EGLNIGDSMGVEODLLVRLLSAAVCDPLGITGYKAKTALGEHLNVLNVGRDNVSEILQI 986
Db      745 EGLICQGSLLDKVQDLLVRLKALHDPGLPPYQCSLKILGEKMESEIPLTRDNVSEILRC 804
Qy      987 FMEAHCGQGTELTSKTKAFOAHTPAQKASVLAFLINELACSKSVVSEIDKNIDYMSNR 1046
Db      805 FLMAVRVPSFCDLSLRTQFOAQPQQAAILAFLVHLSNSTIINIBDKTLESVSSCR 864
Qy      1047 RDKWVEGKIRKLRIIHAHKTGKRD--SGGIDGEEOHPLGTPPGRKRRKGGSDYD 1104
Db      865 KNWIVEGLRLRLTALAKRTGRPEVMMEGAB-----GLGRRSSRIME-- 909
Qy      1105 DDDDDDDDDQDEDEDEDEK---EDOKGKKTIDCEDEDEGQAA-SVEELEKQLEKLSK 1160
Db      910 -----ETSGIEEBEENTAVHGRGRK-----EGEIDVAASSIPELHERHIEKLSK 956
Qy      1161 QOSQYRKLFDASHLSRSMFGRYRRYRWWILPRCGGIFVEGMESGEGLEIAKEREKL 1220
Db      957 RQLFRKLLHSSQMLRAVSLQDQYRRYRHWVLPYLAGIFVEGSE-----1001
Qy      1221 KKAESVQIKEEMFETSGDSLNCSTHCEQEDLKEKDNLTNLFQKPGSFSLKLEVA 1280
Db      1002 -----1001
Qy      1281 KMPPESEVMTKPNAGANGCTLSYQNSGKHSIGSVQSTATQSNVEKADSNLNTGSSGP 1340
Db      1002 -----GSTGP 1006
Qy      1341 GKFSPLNDQLLTKLTEKNQWFSLLPRTPCDDTSLTHADMTASLTVPQSPKSPS 1400
Db      1007 -----WFNESAQIPCD-----AAPTTPPAVSEDO 1030
Qy      1401 PTPA-----PLGS-SAQNPVGLNPALSPQLVKGVSMMGLQFCGWPVTGVTNTPF 1451
Db      1031 PTPSLQLLASSKPMNTPGAANP--CSPVQLSSTHLPGGT-----PK 1069
Qy      1452 TSVPSLSGLGLSEGNGNSFLTNSVASSKESVPQNEKATSQAAPAAVEAKVPDFF-- 1509
Db      1070 RLS-----GDS-----EEMSQSPGLQPKRGRPPS-KFFKQVEQHYL 1107
Qy      1510 ---SPKPIPEMQFGWRIIDPEDLKALLKVLHLRGIREKALQKIQKHLDYITQACLKN 1566
Db      1108 TQLTAQPIPPMCSGWWIWRDPETLDVLLKALHPRGIREKALHKLHLSKHDFLQVCLQ 1167
Qy      1567 KDVAIIELNE---NEENQVTRDIVENKSVESQAMMDLSVLQOVEDLERRVASASLQVG 1623
Db      1168 LTDPFEPNELPALEEG-----VMSWSPKEKTYETDLAVLQWVELEQRVVLSDQIRG 1221

```

```
Qy 1624 WMCPEASERDVLVFEHKSFTLCKEHDGFTGDESSAHALERKSDNPLDIAVTRLAD 1683
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1222 WTCPTDSTREDLTycehLpDS----PEDIWRGREGT-VPOQNNPLDLAVMLAV 1276
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1684 LERNIER-----IEBDIAPGLRVRRALSAR 1711
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1277 LEQNVERRYREPLWAAHVVEKALLSTPNAGPDGTSTSEIYEITPRVVRQTLRCR 1336
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1712 SAAQVALCIGLOKSIANEKIMKYCOICRKGNEELLLLCDGCKGCHTYCHRPKIT 1771
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1337 SAAQVCLCWGLERSIAEKSVNKTCLVCRKGNDEFLLLCDGCDRGCHTYCHRPKME 1396
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1772 IPDGMWFCACIATASGQTLKILHVKKTNESKKGKVTLT---GDTEDDSASTSS 1828
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1397 VPEGMWFCVCLSVQVEEYTORPGPKGQKXSS-----PFLTFPEGDS----- 1441
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1829 SLKRGNKDLOKRM-----EENTSNLSKQESFTSVKKPK---RDDSKDLALCSMLTME 1881
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1442 -----RRMLSRSDSPAVPRYPEDGLSPKRRHSMRSHSHSLTCEIILMEME 1491
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1882 THEDAMPPLPNLKVGYKKVKKPMDFSTIREKLSGQYPNLETALDVLRLVFDNCE 1941
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1492 SHDAWPELEFVNPRLVSGYRVIKNPMDFSTMRELRLLRGYTSSEFAADALLVFDNCQ 1551
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1942 TFNEDSDIGRAGHNMRYFEKKWTDTEK 1970
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1552 TFNEDDSVKGAGHVMRRFFESRWEFYQ 1580
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 15
US-09-864-761-33808
; Sequence 33808, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wenheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aeoimica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
```

```
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 33808
; LENGTH: 194
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC008277.1
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2.3
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.7
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.2
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.94
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.2
; OTHER INFORMATION: EST HUMAN HIT: T19315.1, EVALUE 1.00e-52
; OTHER INFORMATION: SWISSPROT HIT: P53349, EVALUE 3.70e-01
US-09-864-761-33808
```

```
Query Match 10.1%; Score 1021; DB 9; Length 194;
Best Local Similarity 100.0%; Pred. No. 4e-42;
Matches 194; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1232 MFETSGDSLNCSTNDHCEQKEDLKEKONTNLFLOKPGSFSLKLLVAKMPPSEVWTP 1291
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1 MFETSGDSLNCSTNDHCEQKEDLKEKONTNLFLOKPGSFSLKLLVAKMPPSEVWTP 60
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 1292 KPNAGANGCTLSYONSGKHSIGSVQSTATOSNVKADSNLNFNTGSSGPGKFYSLPLNDQ 1351
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 61 KPNAGANGCTLSYONSGKHSIGSVQSTATOSNVKADSNLNFNTGSSGPGKFYSLPLNDQ 120
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 1352 LLKLTLEKNRQWFSLLPRTPCDDTSLTHADMSTASLVTPQSPFSPKSPPTPAPLGSSAQ 1411
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 121 LLKLTLEKNRQWFSLLPRTPCDDTSLTHADMSTASLVTPQSPFSPKSPPTPAPLGSSAQ 180
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 1412 NPVGLNPFALSPLO 1425
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 181 NPVGLNPFALSPLO 194
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
RESULT 16
US-09-839-479-65
; Sequence 65, Application US/09839479
; Publication No. US20020039779A1
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042002
; CURRENT APPLICATION NUMBER: US/09/839,479
; CURRENT FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 65
; LENGTH: 176
; TYPE: PRT
```


ORGANISM: Homo sapiens
US-09-839-479-65

Query Match
Best Local Similarity 8.8%; Score 888; DB 9; Length 176;
Matches 176; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 850 LEORLELEMAKELKKPNEDMCLADQKPLPELPRIPLGLVLSGTSFSDCLMVVQFLRNFGK 909
DB 1 LEORLELEMAKELKKPNEDMCLADQKPLPELPRIPLGLVLSGTSFSDCLMVVQFLRNFGK 60

QY 910 VLGFVDNIDVNLVQLQGLNIGDSMGEVDLLVRLLSAAVCDPGLITGYKAKTALGEH 969
DB 61 VLGFVDNIDVNLVQLQGLNIGDSMGEVDLLVRLLSAAVCDPGLITGYKAKTALGEH 120

QY 970 LLNVGVNRDNDVSEILQIFMEAHCGQTTELTSKTKFAQHTPAQKASVLAFLINEL 1025
DB 121 LLNVGVNRDNDVSEILQIFMEAHCGQTTELTSKTKFAQHTPAQKASVLAFLINEL 176

RESULT 17
US-10-537-66

Sequence 66, Application US/10376537
Publication No. US20030224405A1
GENERAL INFORMATION:
APPLICANT: Jones, Michael H.
TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
CURRENT APPLICATION NUMBER: US/10/376,537
CURRENT FILING DATE: 2003-02-28
PRIOR APPLICATION NUMBER: US/09/418,710
PRIOR FILING DATE: 1999-10-15
PRIOR APPLICATION NUMBER: PCT/JP98/01783
PRIOR FILING DATE: 1998-04-17
PRIOR APPLICATION NUMBER: JP 9/310027
PRIOR FILING DATE: 1997-10-24
PRIOR APPLICATION NUMBER: JP 9/116570
PRIOR FILING DATE: 1997-04-18
NUMBER OF SEQ ID NOS: 73
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 66
LENGTH: 176
TYPE: PRT
ORGANISM: Homo sapiens
US-10-376-537-66

Query Match
Best Local Similarity 8.8%; Score 888; DB 15; Length 176;
Matches 176; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 850 LEORLELEMAKELKKPNEDMCLADQKPLPELPRIPLGLVLSGTSFSDCLMVVQFLRNFGK 909
DB 1 LEORLELEMAKELKKPNEDMCLADQKPLPELPRIPLGLVLSGTSFSDCLMVVQFLRNFGK 60

QY 910 VLGFVDNIDVNLVQLQGLNIGDSMGEVDLLVRLLSAAVCDPGLITGYKAKTALGEH 969
DB 61 VLGFVDNIDVNLVQLQGLNIGDSMGEVDLLVRLLSAAVCDPGLITGYKAKTALGEH 120

QY 970 LLNVGVNRDNDVSEILQIFMEAHCGQTTELTSKTKFAQHTPAQKASVLAFLINEL 1025
DB 121 LLNVGVNRDNDVSEILQIFMEAHCGQTTELTSKTKFAQHTPAQKASVLAFLINEL 176

RESULT 18
US-10-702-148-65

Sequence 65, Application US/10702148
Publication No. US20040063145A1
GENERAL INFORMATION:
APPLICANT: Jones, Michael H.
TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
CURRENT APPLICATION NUMBER: US/10/702,148
CURRENT FILING DATE: 2003-11-05

PRIOR APPLICATION NUMBER: PRIOR APPLICATION NUMBER: US/09/839,479
PRIOR FILING DATE: 2001-04-20
PRIOR APPLICATION NUMBER: US 09/418,710
PRIOR FILING DATE: 1999-10-15
PRIOR APPLICATION NUMBER: PCT/JP98/01783
PRIOR FILING DATE: 1998-04-17
PRIOR APPLICATION NUMBER: JP 9/310027
PRIOR FILING DATE: 1997-10-24
PRIOR APPLICATION NUMBER: JP 9/116570
PRIOR FILING DATE: 1997-04-18
NUMBER OF SEQ ID NOS: 72
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 65
LENGTH: 176
TYPE: PRT
ORGANISM: Homo sapiens
US-10-702-148-65

Query Match
Best Local Similarity 8.8%; Score 888; DB 15; Length 176;
Matches 176; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 850 LEORLELEMAKELKKPNEDMCLADQKPLPELPRIPLGLVLSGTSFSDCLMVVQFLRNFGK 909
DB 1 LEORLELEMAKELKKPNEDMCLADQKPLPELPRIPLGLVLSGTSFSDCLMVVQFLRNFGK 60

QY 910 VLGFVDNIDVNLVQLQGLNIGDSMGEVDLLVRLLSAAVCDPGLITGYKAKTALGEH 969
DB 61 VLGFVDNIDVNLVQLQGLNIGDSMGEVDLLVRLLSAAVCDPGLITGYKAKTALGEH 120

QY 970 LLNVGVNRDNDVSEILQIFMEAHCGQTTELTSKTKFAQHTPAQKASVLAFLINEL 1025
DB 121 LLNVGVNRDNDVSEILQIFMEAHCGQTTELTSKTKFAQHTPAQKASVLAFLINEL 176

RESULT 19
US-09-764-864-1165

Sequence 1165, Application US/09764864
Patent No. US20020132753A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PT223
CURRENT APPLICATION NUMBER: US/09/764,864
CURRENT FILING DATE: 2001-01-27
Prior-application data removed - consult PALM or file wrapper
NUMBER OF SEQ ID NOS: 1792
SOFTWARE: Patent in Vet 2/0
SEQ ID NO 1165
LENGTH: 160
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (2)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-864-1165

Query Match
Best Local Similarity 8.1%; Score 817; DB 9; Length 160;
Matches 156; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1815 TGDTEDEDSASTSSSLKRGNDLQKCKMEENTSNLSKQESTSVKKPKRDDSKDALCS 1874
DB 3 TGDTEDEDSASTSSSLKRGNDLQKCKMEENTSNLSKQESTSVKKPKRDDSKDALCS 62

QY 1875 MILTEMETHEDAWPFLPVLNKLVPYKVKIKKPMDFSTIREKLSGGQYPNLETFALDVR 1934
DB 63 MILTEMETHEDAWPFLPVLNKLVPYKVKIKKPMDFSTIREKLSGGQYPNLETFALDVR 122

QY 1935 LVFDCNCFNEDDSDIGRAGHNMRYKFEKKWTDTPKVS 1972
DB 123 LVFDCNCFNEDDSDIGRAGHNMRYKFEKKWTDTPKVS 160

RESULT 20

US-09-839-479-1

; Sequence 1, Application US/09839479

; Publication No. US20020039779A1

; GENERAL INFORMATION:

; APPLICANT: Jones, Michael H.

; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR

; FILE REFERENCE: 06501-042002

; CURRENT APPLICATION NUMBER: US/09/839,479

; CURRENT FILING DATE: 2001-04-20

; PRIOR APPLICATION NUMBER: US 09/418,710

; PRIOR FILING DATE: 1999-10-15

; PRIOR APPLICATION NUMBER: PCT/JP98/01783

; PRIOR FILING DATE: 1998-04-17

; PRIOR APPLICATION NUMBER: JP 9/310027

; PRIOR FILING DATE: 1997-10-24

; PRIOR APPLICATION NUMBER: JP 9/116570

; PRIOR FILING DATE: 1997-04-18

; NUMBER OF SEQ ID NOS: 72

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 1

; LENGTH: 1674

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-839-479-1

Query Match

Best Local Similarity 21.5%; Score 726.5; DB 9; Length 1674;

Matches 338; Conservative 186; Mismatches 424; Indels 627; Gaps 53;

Qy	664	SRMRKCPNPVGNNAEFLDNADAKLRKLAQAEIARQAQIKLRKLKQEQEQAARVAKA	723
Db	420	SPANRRGRPP-----KRIHSQDNVA---442	
Qy	724	KQQAIAAEKRRKQEKQIKIMQOEKIKRIQIRMEKELRAQQLLEAKKKKEEAANAK	783
Db	443	--NQTLASYRSKATKEDKLLKQEE---MKSLAFE-----KAKLKREKADA-484	
Qy	784	LLEAKRIKEKEMRRQQAALLKQERERRRRHMMKAMKAEKAKERLQKQRDEKR	843
Db	485	-LEAKK--KEKE-----DKEKKRBEKLVVEERLKKKEEKERLKVEREKE--527	
Qy	844	LNKERKLEQRLEEMAKELKPNEDMCLAQKPLPELPRIPLGLVLSGTSFSDCLMVQF	903
Db	528	--REKLEEKRYVEYLKQWSKPRDMECDLKLPE-PTPVKTRLPPEIFGDALMVLEF584	
Qy	904	LRNFGKVLGFVNIDVFN---LSVLQGLLNIIGDSMGVQDILLVRLLSA---AVCDPGLI957	
Db	585	LNAFGL--FDLQDEFPGVTLVLEALVG-NDESGPLCELLPFFLTATFQATAEBEE641	
Qy	958	TGYKAKT---ALGEHLNVGNRDNVSILOIF-----987	
Db	642	VAKEQLTDATKGCGLSLDSDCTLSILRLHILASGADVTSANAKRYQKRGFGDATD701	
Qy	988	---MEAHCGQETELTESLTKFAQHTPAQKASVLAFLINELACSKSVSEIDKNIIDYMSN1044	
Db	702	DACMELRLNSPLVKJLSSTSYDILTPEGKKILHALCGKLL---TLVSTRDFTEDYDI758	
Qy	1045	LRRDKWV-----EGKLR-----KLRILHAKTKGRDTSGG1075	
Db	759	LRAQKQEFRELKAEQHRKERBEAARIRKKEELKEQEQKKEKQELKEDQBNSTAD818	
Qy	1076	IDLGEQHPGLGTPTPGRKRRKGGSDYDDDDDDDDQDQDEDEED----KEDQKKG1131	
Db	819	LSIGEE-----REDPDTIESKQTEQKELDQDMFTEDEDDPGSHKGRGK865	
Qy	1132	KTD-----ICDEDEGQAASVELEKQIEKLKQOQYRRKLPDA1172	
Db	866	RGQNGFKETRQEQINCVTRELLTADEEE---ALKQEHQKEKELLEKIQS-----A914	
Qy	1173	SHSLRSVMFGPDYRRRRYWLPRCGGIFVEGMEGEGLEETAKEREKLUKAEVQIKEM1232	

Db	915	IACNIPPLGRDRMYRRYWIPIPSIPGLFIE--EDYSGLTE-----952	
Qy	1233	FETSGDSLNCNTSDHCQKEDLKEKDNNTNLFLOKPGSFSKLKLEVAKMPESEVMTPK1292	
Db	953	-----DMLLPPRSFQN-----NVQSQDPQVSTKTGE979	
Qy	1293	PNAGANGCTLSYQNSKGHSLGSVOSTATQSNVEKADSNLFTGSGGPGKFVPLPND--1350	
Db	980	P-----IMSEST-SNIDOG-----PRDHS997	
Qy	1351	-QLLKLTLEKRNWFSLLPRTPCDDTSLTHADMSTASLVTQSQPPSPSPPTAPLGSS1409	
Db	998	VOLPKPVHKPNR-W-----CFVSSCEQLDQLEAL-----1026	
Qy	1410	AQNPVGLNPPFALSPLQVKGGVSMGLQFCGWTGVTSTNIPFTLSVPSLSGLGLSEGNG1469	
Db	1027	-----NSRGR1032	
Qy	1470	NSFLTSNVASSKSESPPVQNEKATSAQPAAVEVAK--PVDPPSPKPIPEEQFGWRIID1527	
Db	1033	ESALKETILLQEKSR-----ICAQLARFSEKFFHSDKPDQDSKFTYGRSSNAYD1083	
Qy	1528	PEDLKALLKVLHLR-----GIREKALQ-----KQIQKHLDYITQACLKNKDVAILLN1575	
Db	1084	PSQWCA-EKQLELRDLDFLLDIEDRIYQGLGAIKVTDRH--IWRSALESGRVELLS-E1138	
Qy	1576	ENENQVTRDIVENWSVEEQAMENDLSVLQOVDELERRVASASLOVKGWMCPEPASERED1635	
Db	1139	ENKENGIIKTV--NEDVEEMBEDQTKVIVKDRLLGKTKETPSTVSTNASTPQSVS----1192	
Qy	1636	LVYFEHKSFTKLCKEHGDEFTGEDESSAHALERKSDNPLDIAVTRLADLENIERRIBED1695	
Db	1193	-----SVVHYL-----AMALFOIEGIERFLK-1215	
Qy	1696	IAP-----GLRVRRALSEARSAQVALCIQLOLQSAWEKSIIMKYVCQICRK1743	
Db	1216	-APLDASDGRSYKTVLDRWRESLLSSASLSQVFLHLSTLDRSVIWSKSIILNARCKICRK1274	
Qy	1744	GDNEELLLLCDGCKGCHTYCHRPKITIPGDWFCACIAKASQTLIKIKLHVKGKKT1803	
Db	1275	KGDAENMVLCDGCDRHHTYCVRPKLKTVPEGDWFCPCRPKQ-----RCRRLSFRQRP1329	
Qy	1804	NEKKK-----GKVTTLTGDT-----DED-----1822	
Db	1330	LESDVEDSMGGDDVDGDEEGQSEEEYVEQDEDDSQSEEEVSLPKRGRPQVRLP1389	
Qy	1823	-----SASTSSSL-----1830	
Db	1390	VKTRGKLSSPSRGOQEPGRYPSPRSQSQSTPKTTVSSKTSRSLRKINSAPPTETKSLRI1449	
Qy	1831	-----KRGKD-----LQKRMKEENTSI1848	
Db	1450	ASRSTRSHGFLQADVVELLSPRKRRGRKSANNTPENSPFPNFVIAATKSSEQSRSV1509	
Qy	1849	N-----LSQESFTSVKPKRDDS------DLALCSMLITEMETHDA1886	
Db	1510	NIASKLSQSESEKRRCKRKQSPSPVTLGRSSRGQGVHLSAFQLVVLVRHDDS1569	
Qy	1887	WPFLLPVNLKLVPGYKVKIKPMDFTSIREKLSGQYPNLFTFALDVLVFDNCFTFED1946	
Db	1570	WPFLKLVSKIQVPDYDIKKPIALNIREKVNCKEYKLASEFIDDDIELMFNCFEYNPR1629	
Qy	1947	DSDIGRAGHNWKYF1961	
Db	1630	NTSEAKATRLQAPF1644	

RESULT 21

US-10-376-537-1

; Sequence 1, Application US/10376537

; Publication No. US20030224405A1

; GENERAL INFORMATION:


```

; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 69
; LENGTH: 1673
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-839-479-69

Query Match          7.0%; Score 709; DB 9; Length 1673;
Best Local Similarity 21.0%; Pred. No. 7e-26;
Matches 329; Conservative 204; Mismatches 426; Indels 604; Gaps 51;

Qy 664  SPRRRRKGRPNVGNAAEFLNADAKLRKLAQAEIARQAAQIKLRKLQEQEQAIVAKEA 723
Db      |||:||||
420  SPANRRRRGRPP-----KRIHISQEDNVA--- 442

Qy 724  KKOQAIWAABEKRKQEQIKIMKQOEKI KRIQOIRMEKELRAQILAEAKKKKEEANA 783
Db      |||:||||
443  --NQOTLASVRSKATKRDKLLQEE-----MKSIAFE-----KAKLRKADA- 484

Qy 784  LLEAKRKIKBKEMRRQQAIVLLKHQERRRRQHMLMKAMARKAKBEKERLKEKRDKR 843
Db      |||:||||
485  -LEAKK--KEKE-----DKEKREELKKIVVEERLKKCEKERLKYEREKE-- 527

Qy 844  LNKERKLEQRLEMAKELKKPNEDMCLADQKPLPELPRIGLVLVSGSTFSCLMVQF 903
Db      |||:||||
528  --REKUREEKKKVEY--KOWSKREDMECDLDELPE-PTPVKTRLPPEIFGDALMVLEF 583

Qy 904  LRNFQKVLGPDVNIQVNP--LSVLQEGILNIGDSMGVEODLLVRLISA--AVCDPGLI 957
Db      |||:||||
584  LNAFGEI--FDLOQEPDGVTVLEBALVG-NDSEGPICELLPFFLTALFOAIAEEEE 640

Qy 958  TGYKAKT--ALGEHLNVGNRDNVSEILQIF----- 987
Db      |||:||||
641  VAXEQLTADATKGCGLSKLDLSDCTLSEILRLHLASGADVTSANAKRYOKRGGFDA 700

Qy 988  ---WEAHCGQTELTESLTKTAFQAHPAQAQASVLAFILINELACSKSVVSEIDKNIDYMS 1044
Db      |||:||||
701  DACMEURLSNPSLVKKLSSTSVVDLTPEGKKVILHALCGKLL---TLVSTRDFIEDYVDI 757

Qy 1045  LRREDKVV-----EGKLR-----KLRILHAKTKTKRDTSGG 1075
Db      |||:||||
758  LRQAKQEFRELKAEQHKEREEREAARIRKEKEKLEQEQKMKKEQKLEKEDQRNSTAD 817

Qy 1076  IDLGEQHPGLGTPTPGRKRRRKGGSDYDDDDDDDDQDEDEDEDEDEDEDEDEDEDEDE 1131
Db      |||:||||
818  ISIGEE-----REDFDTIESKTEQKEQLDQDMFTEDDDPGSHKRRGRGK 864

Qy 1132  KTD-----ICEDEGDOAASVEELEKQLEKLSKOOSQYRKLFDA 1172
Db      |||:||||
865  RGQNGFKFTRQGINCVTRRELTADEEE---ALKQEHQKKEKLELLEKLSQ-----A 913

Qy 1173  SHSLRSVMFDPDRYRRYRWILPRCGGIFVBEQMSSEGLEIRIAKEREKLLKAEASVQIK 1232
Db      |||:||||
914  IACTNIFPLGRDWRYYRWIFPSLPGLFIE--EDYSGLTEDHLLPRSSFQNNVQSDPQ 971

Qy 1233  FET--SGDSLNCSTNDHCEQKEDLKEKDNWTLFQKPGSFKSLKLEVAKMPPESEVMTP 1291
Db      |||:||||
972  VSTKTEGLMSESTNIDQG-----PRDHSVQLP 1000

Qy 1292  KPNAGANGCTLSYONSGKHSVQSTATQSNVEKADSNLNTVGTSGGPK--FYSPLPN 1349
Db      |||:||||
1001  KP-----VHK-----PNRWCFYSSCEQ 1017

Qy 1350  -DQLLTKLTETKNQWFSLLPRTPCDDTSLTHADMSTAS----LVTPQSOPPSK----- 1397
Db      |||:||||
1018  LDQLLEALNSRGRH--ESALKETILQEKSRICAQLARFSEEBKFHESDKRPDSKPTYSRGR 1076

Qy 1398  -SPSPPTPAPLGGSAQNPVGLNPNPALSPQLVQGVGVMGLQFCGWPTGVVTSNTPFTLSVP 1456
Db      |||:||||
1077  SSNAYDPSCABKQLELRDLFLL-----DIEDRIYQG 1110

```

Qy	1457	SLSGGLGLSGNGNSFLTSTNVASSKSGESPVPQNEKATSAQPAAVEVAKPVDFFSPKPIPE	1515
Db	1111	TLG	1125
Qy	1517	EMQFGWRIIDPEDLKALLVHLRGIREKALOKIQKHLDYITQACLKNKDVALLLENE	1576
Db	1126	ALBSGRYELLUSEN	1159
Qy	1577	NEENQVTRDITVENNSVEEQAMEMDLSVLQOVEDLERRVVASASLQVKGWMCPEPASBERDL	1636
Db	1160	QTKVIVKDRLLGIKTETPTSTVSTNASTPQSVSSVVHVALWALQJIE	1205
Qy	1637	VYFEHKSFTKLCKEHDEFTGEDESSAHALERK-SNPPLDIATRIADLERNTERIEED	1695
Db	1206	QGLERRFLKAPLD	1232
Qy	1696	IAPCLRVRREALSEARSAAQVALCIOOLOKSIWAKSIMKVYCOICRKGDNELLALLCDG	1755
Db	1233	WRESLLSSASLSQVFLHLSTLDKSVTWSKSIILNARCKICRKGDAENMVLLCDG	1285
Qy	1756	CDRGCHTYCHRPKITTPDGDWFCPACIAKASQOTLKIKKLHVKGKKTINESK-	1808
Db	1286	CDRGHHTYCVRPKLKIVPEGDFCPECRPKQ	1340
Qy	1809	GKKVTITLGDTE	1822
Db	1341	GEDDEVDGBEEQSGBEEYEVEQDEDDSQEBEVSILPKRGRPQVRLPVKTRGKLSSSF	1400
Qy	1823	SASTSSSL	1930
Db	1401	SRGOQEGPYRSRSQOSTPTVTTSVSKTGRSLRKINSAPPTETKSIARIASRSTRHSHGPL	1460
Qy	1831	XRGNKD	1856
Db	1461	QADVVELLSPRKRGRKRSANNTPENSPNFPNFRVIATKSSSQSRVNIASKLSQESE	1520
Qy	1857	TSVKPKRDDSK	1898
Db	1521	SKRCKRQSPPEPSPVTLGRSSRGQGVHELSAFEQLVVELVRHDDSWPFLKLVSKIQV	1580
Qy	1899	PGYKVKTKPMDFSTIREKLSGQYFNLETFALDVLRLVDCNCFNEDDSDDIGRAGHNR	1958
Db	1581	PDYDIIKKPIALNIREKNVCKEYKLASEFIDDIELMFNSCNCFEYNPRNTSEAKAGTRLQ	1640
Qy	1959	KYF	1961
Db	1641	APF	1643
RESULT 24			
US-10-376-537-70			
; Sequence 70, Application US/10376537			
; Publication No. US20030224405A1			
; GENERAL INFORMATION:			
; APPLICANT: Jones, Michael H.			
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR			
; FILE REFERENCE: 06501-042001			
; CURRENT APPLICATION NUMBER: US/10/376,537			
; CURRENT FILING DATE: 2003-02-28			
; PRIOR APPLICATION NUMBER: US/09/418,710			
; PRIOR FILING DATE: 1999-10-15			
; PRIOR APPLICATION NUMBER: PCT/JP98/01783			
; PRIOR FILING DATE: 1998-04-17			
; PRIOR APPLICATION NUMBER: JP 9/310027			
; PRIOR FILING DATE: 1997-10-24			
; PRIOR APPLICATION NUMBER: JP 9/116570			
; PRIOR FILING DATE: 1997-04-18			
; NUMBER OF SEQ ID NOS: 73			
; SOFTWARE: Fast-Seq for Windows Version 4.0			
; SEQ ID NO 70			
; LENGTH: 1673			
; TYPE: PR			
; ORGANISM: Homo sapiens			

US-10-376-537-70

Query Match 7.0%; Score 709; DB 15; Length 1673;
Best Local Similarity 21.0%; Pred. No. 7e-26;
Matches 329; Conservative 204; Mismatches 426; Indels 604; Gaps 51;

QY 664 SRMRKRGRPPNVGNAEFLDNADAKLRLKLAQAEIARQAQKILRLKQKQOARVAKEA 723
DB 420 SPANRRGRPP-----KRIHISQEDNVA--- 442

QY 724 KQQAIAAAEKKRQKEQIKIMQOQEKIKRIQOIRMEKELRAQOILEAKKKKEEANAANAK 783
DB 443 --NKQTLASYSKATKRDKLLKQEE---NKSIAFE-----KAKLKREKADA- 484

QY 784 LLEAKRIKEKEMRROQAAILKHQERRRRQHMMLKMAEAKAABEKERLQKQRDKR 843
DB 485 -LEAKK--KEKE-----DKEKKREBELKIVEEERLKKKEERLKVREKE-- 527

QY 844 LNKERKLEORLEMAKELKPNEDMCLADQKPLPELPRIPGLVLSGSTFSDCLMVVOF 903
DB 528 --REKLREKKYBY-KQWSPREDMCEDDLKELPE-PTPVKTRLPPEIFGDALMVLEF 583

QY 904 LRNFQVLGFDVNDVFN---LSVLQEGLLNIGSMGEVQDILLVRLLSA---AVCDPGLI 957
DB 584 LNAFGEI--FDLQDEFPDGVTLVELEALVG-NDSEGPLCELLFFFLTAIFAQIAEBEEE 640

QY 958 TGYAKT--ALGEHLNVGNRDNVSEILQIF----- 987

DB 641 VAKEQLTDADTKGCSLKSLLDSCSLSEILRLHLILASGADVTSANAKRYQKRGGFADT 700

QY 988 ---MEAHCGQETELSTKTAFOAHTPAKASVLAFINELACSVVSEIDKNIDYMSN 1044
DB 701 DACMELRSLNPLVKLSSTSYDUTPGEKWKILHALCGKLL---TLVSTRDFIEDYDI 757

QY 1045 LRRDKWV-----EGKLR-----KLRIHAKKTGKRDTSGG 1075
DB 758 LRQAQEFRELKAEQHRKEREEMAAIRKRKEELKEQEQMKKEQKLEKEDQERNSTAD 817

QY 1076 IDLGEQHPGTPTPRKRRKRGSDYDDDDDDDDQDGEDDEDED---KEDQKQK 1131
DB 818 ISIGEE-----REDFOTSIESKDTEQKELDQDMFTEDDDPGSHKRRGRGK 864

QY 1132 KTD-----ICDEDEGQAAASVELEKQIEKLSKQSQOYRRKLFPDA 1172
DB 865 RGQNGFKFTQEQINCVTRELLTADEE---ALKQEHQREKELLEKLQS-----A 913

QY 1173 SHLSRVNMGDPDRYRRYWIPLRCGGIFVEGMESGEGLEEIAKEREKLKKAESVQIKBEEM 1232
DB 914 IACTNIFFPLGRDRMYRWIPLPLGLFIE--EDYSGLTEDHLLPRPSSFQNNVQSQDPQ 971

QY 1233 FET-SGDSLNCNTHCQEKEDLKEKONTNLFQKPGSFSLKLELVAKMPPSEVWTP 1291
DB 972 VSTKTGEPLMSESTNIDQG-----PRDHSVQLP 1000

QY 1292 KPNAGANGCTLSYQNSGKHLSGVSTATQSNVEKADSNLNTGSSGPGK--FVSLPLN 1349
DB 1001 KP-----VHK-----PNRWCFFYSCEQ 1017

QY 1350 -DQLLTLTEKNRQWFSLLRTPCDDTSLUTHADMSTAS---LVTPSQQPPSK----- 1397
DB 1018 LDQILEALNSRGR--ESALKETLLQEKSRICAQLARFSEKHFHFDKQPDSPKTYSGR 1076

QY 1398 -SPSTPAPLSSAQNPGVLPFALSPLQVKGVSMMGLQFCGWPVTGVVTSNIPFTLSVP 1456
DB 1077 SSNAYDPSQMAEKQLELRDLRDLFLL-----DIEDRIYQG 1110

QY 1457 SIGSLGLSGNGNSFLTSNVASSKSPVPQNEKATSAQPAAVEKVPDPFPKPIPE 1516
DB 1111 TLG-----AIKVTDRHIWRS----- 1125

QY 1517 EMQFGWRIIDPEDLKALLKVLHLAGIREKALQKIQKHLDYITQACLKKNKDVAIIELE 1576
DB 1126 ALESGRYELLSEEN-----KENGIIKTV-----NEDVEWEIDE 1159

QY 1577 NEENQVTRDIVENWSVEEQAMEMDLVLIQQVEDLERRVVASASLOVKGWMCPEPASERDL 1636
DB 1160 QTKVIVKDRLLGIKITEPTSTVSTNASTPQSVSSVVHYLMAALFOIE----- 1205

QY 1637 VYFEHKSFTKLCKEHDGEFTGEDESSAHALERK--SDNPLDIAVTRLADLERNIRRIEED 1695
DB 1206 -----OGLERRFLKAPLD-----ASDSGRSYKTVLDR- 1232

QY 1696 IAPGLRWRRLSAPRSAAQVALCICQLOKXIAWEKIMKVYCOICRKGDNELLLLCDG 1755
DB 1233 -----WRESLLSSASLSQVFLHLSTLDRSVIWSKILNARCKICRKGDAENMVLCDG 1285

QY 1756 CDKCHFYCHRPKITITPDGDMFCPACIAKASGOTLAKLHVKGKKTNESKK----- 1808
DB 1286 CDRGHTYCVRPKLKIVPEGDMFCPCRPKQ-----RCRRLSFRQRPSEDESDEVEDSMG 1340

QY 1809 GKVTYLTGDTY-----DED----- 1822
DB 1341 GEDDEVGDDEEGQSEEEVEVEQEDDSQEEEEVSLPKRCRPQVRLPVKTRGKLSSESFS 1400

QY 1823 -----SASTSSSL----- 1830

DB 1401 SRGOQBPGRYPSPRSQOSTPKTTVSSKTGRSLRKINSAPPTETKSLRIASRSTRHSGPL 1460

QY 1831 -----KRGNKD-----LQKRWKEENTSIN---LSKQESF 1856

DB 1461 QADVVELLSPPRRKRRGRKGSANNTPENSPFPNPRVIATKSSQESRSVNIASKLSQSE 1520

QY 1857 TSVKKPKRDRDSK-----DLALCSMLITEMETHEDAWPFLPVNLKLV 1898

DB 1521 SKRCRKRQSPSPVTLGRRSSRGQGVHLSAFEQLVVELVRHDDSWPFLKLSKIQV 1580

QY 1999 PGYKVKIKKPMDFSTIREKSSGOYPNLETFALDVLVLFVFNDCETTFNEDSDSIGRAGNMNR 1958
DB 1581 PDYDIYIKKPIALNIIREKVNKCYKLASEFIDDIELMFNSCFEYNPNTSEAKAGTRLQ 1640

QY 1959 KYF 1961
DB 1641 AFF 1643

RESULT 25
US-10-702-148-69
; Sequence 69, Application US/10702148
; Publication No. US20040063145A1
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042002
; CURRENT APPLICATION NUMBER: US/10/702,148
; PRIOR FILING DATE: 2003-11-05
; PRIOR APPLICATION NUMBER: PRIOR APPLICATION NUMBER: US/09/839,479
; PRIOR FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 69
; LENGTH: 1673
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-702-148-69

Query Match 7.0%; Score 709; DB 15; Length 1673;
Best Local Similarity 21.0%; Pred. No. 7e-26;
Matches 329; Conservative 204; Mismatches 426; Indels 604; Gaps 51;

NAME/KEY: SITE
LOCATION: (130)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-864-1578

Query Match 5.7%; Score 582; DB 9; Length 134;
Best Local Similarity 91.9%; Pred. No. 5.2e-21;
Matches 114; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

1831 KGNKDLQKRWMEENTSINLSKQESFTSVKKPKRDDSKDLALCSMILTEMETHEDAWPFL 1890
1 KGNKDLKCRKMEENTSINLSKQESFTSVKKPKRDDSKDLALCSMILTEMETHEDAWPFL 60
1891 LPVNKLVPYGVKVIKKPMDFTIREKLSGGQYPNLETFALDVLVDFPNCETENEDDSDI 1950
61 LPVNKLVPYGVKVIKKPMDFTIREKLSXGQYPNLETFALXVLVDFTFMCKXXDI 120
1951 GRAG 1954
121 GRAG 124

RESULT 27

US-09-839-479-68
Sequence 68, Application US/09839479
Publication No. US20020039779A1
GENERAL INFORMATION:
APPLICANT: Jones, Michael H.
TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
FILE REFERENCE: 06501-042002
CURRENT APPLICATION NUMBER: US/09/839,479
CURRENT FILING DATE: 2001-04-20
PRIOR APPLICATION NUMBER: US 09/418,710
PRIOR FILING DATE: 1999-10-15
PRIOR APPLICATION NUMBER: PCT/JP98/01783
PRIOR FILING DATE: 1998-04-17
PRIOR APPLICATION NUMBER: JP 9/310027
PRIOR FILING DATE: 1997-10-24
PRIOR APPLICATION NUMBER: JP 9/116570
PRIOR FILING DATE: 1997-04-18
NUMBER OF SEQ ID NOS: 72
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 68
LENGTH: 1525
TYPE: PRT
ORGANISM: Homo sapiens
US-09-839-479-68

Query Match 5.3%; Score 538.5; DB 9; Length 1525;
Best Local Similarity 18.1%; Pred. No. 1.2e-17;
Matches 353; Conservative 253; Mismatches 604; Indels 745; Gaps 64;

189 KPLSLVNOAKKETYMKLI VPSDVLKAGNKNTSESSLLTSELRSKREQY-----KQA 241
8 KPFLVNLPGEEP-FTIPH-----TQEAFTREYEARELERYSERIWTCKST 55
242 FPSQLKQESSKSLKKVLAALSNPKATSSSPA-HPKQTLNNHPNPFLLNGLNQHONG 300
56 GSSQLTHKEAWEEQOEVELLK-----EPPFAWYKGLVLMVHHN-----TA 97
301 VIQSVIQAPLALTTKTKMQSKINENIAAASSTPSSPVNLSTSGRRTPGNQTPWMPAS 360
98 SLEKLVDTAWLEIMTKYAVGECDFEV----- 124
361 PILHSQKKEKAVSNVNPVKTHSHPAKSLVQEPFRGTDSIPSSKQSEDSNEDEEEDDE 420
125 -----GKEKML--KVKIVKI---HPL-----EKVDEE 146
421 EDEDEDDDDSDSQSSEDSNSEDTEGSEEDDDDDKQDESDD-----T 467
147 ATEKSDGACDPSPSDKENSQIAQDHQKCTVVKVEDEGRRESINDRARRSPKPLPTSLK 206
468 EGE-----KTSMKL---NKTTSVKSPSMLTGHSTPRNLHIAK----- 503

Db 207 KGERKWAPPFLPHKYDVKLQNEDKIISNV--PADSLIRTERPPNKEIVRYFIRHNALRA 264
QY 504 -----APGSAPAALCSQSQA-----PLGTSSSTL---TSSPHSGTSKRRAVT 544
Db 265 GTGENAPWVVEDELVKYKSLPSKFSDFLLDPYKMTLNPSTKRNKNTGSPDRKPSKSK-T 323
QY 545 DERELRIPLEYGWQRETRIRNFGRLQGEVAYAPCGKGLROYPEVIKYLNRNGIMDISR 604
Db 324 DNSSLSPL-----NPKLWCHVHLKSL- 347
QY 605 DNFSFAKIRYGDYFYEARDGQEMQWCL-----LKEEDVIPRAMEGRGRPPNPDR 657
Db 348 -----GSPLYKYNKSNKSPPEEHLKEMKMWSPNKLHTNPHIPK-----KGP-PAK 392
QY 658 QRAREESRRRRKGRPNVGNVNAEFLDNADAKLLRLQAOETARQAAQKLLRLKQKQBOA 717
Db 393 KPGKHSKPLKAKGRSGI-----LNGOKST 418
QY 718 RVAKEAKKQQAIAAAEERKQKEOI KIMQOEKIKRIO-----IRMEKEL--RAQ 766
Db 419 GNSKSPKK--GLKTPKTKMKMTLLDMAKGTQKMTAPRNSGGTPTSSSKPHKHLPPAAL 476
QY 767 QILEAKKKKKEE-----AANAKLLEAEKRIK-EKEMR---RQOAVLLKHOER-- 809
Db 477 HLIAYYKENDKRDSALSQVISKRTARLLSSSEDRARLPEELRSLVQKRYELLEHKRWA 536
QY 810 ---ERRQHMLMKAMEARKAEBERKLBQKQDEKLNKERKLBQRRLELEMAKELKPP 866
Db 537 SMSBEQRKEVLLKREELKKLKEKAKERREKEMLEKQRYEQDEL---TGKNL--- 590
QY 867 NEDMCLADQKPLPELPRI PGLVLGSGTFSCLMVVQFLRNFQVGLGVLDVNDVNLVLQ 926
Db 591 -----PAFRLVDTPE--GLPNTLFGDVAMVVEFLSCYSGLLLPDAQPIITAVS-LM 638
QY 927 EGLNIGDSMGEVDLLVRLLSAAVCDPGLITGYKAKTALGHEHLLNVGNRDNVSEIIQI 986
Db 639 EALADKGGFL-YLNRVLVILLQTLQDB-IAEDY---GELGMKLSIEPLTLHSSVSELVRL 693
QY 987 FM-----EAHCGGTETSLKTKAFOAHTPAQKASVLAFLIN 1023
Db 694 CLRRSDVQSESESGDITDDNKDSAAFEDNEVDQEFLEKLETSEFFELTSEELQILTALCH 753
QY 1024 ELACSKSVVSEIDKNIDYMSNLRDKVV---EGKLRKLRI IHAKTGKRTDGGIDLG 1079
Db 754 RILMTYSVQDHMETRQOMSAELWKEKRLAVLKEENDKRAEKQKKEKAKKENGKXWENG 813
QY 1080 EEOHPLGTPTPKRRRRKGGSDYDDDDDDDDDD-----QGDEDEDEDEDE 1126
Db 814 -----LG-----KTDRKRIKVPKPEQVDTAEADMISAVKSRRLLAIOAKKEREIQE--R 860
QY 1127 DQKSKKTDICEDEDEGQAAASVELEKQIEKLSKQSQYRRKLPDASHLSLSVNFQPDY 1186
Db 861 EMVKLERQAEERIRKHAAAEE-----AFQEGIAKAKALVWERTPIGTDRN 907
QY 1187 RRRYWL-PRCGGIFVEGMESGEGLEIAKEREKLLKKAESVQIKKEEFETSG---DSLNC 1242
Db 908 HNRWLFSDVPGFLFIE-----KGWVHDSIDY 934
QY 1243 SNTDHCQKEDLKEKONTNLFQKPGSFKLSKLEVAKMPPSEVMTPKPNAGAGCTL 1302
Db 935 RFNHC-----KDH-----VSGCED 950
QY 1303 SYQNSGKHLSGVOSTATQSNVEKADSNLNTGSSGFGKPYSPLPNDQLLKTUTEKNRQ 1362
Db 951 YCPRSKKANLGKNASMNTQHG-----TATE----- 975
QY 1363 WFSLLPRTPCDDTSLTHADMSTASLVTPSQPPSPSPTPAPLGSSAQNPVGLNPPALS 1422
Db 976 ----- 975
QY 1423 PLQVKGVSMGLQFCGWPTGVVTSNIPFTLSVPSLGLSGLSBGNNGNSFLTSNVASSKS 1482

Db 976 ----- 975
Qy 1483 ESPVONEKATSAQAAPAVEAKVPDFSPKPIPEEMQFGWRIIDPDLKALLKVLHLRG 1542
Db 976 -----VAVETT-----TPKQGNLWFLCDSQKELDELLNCLHPQG 1010
Qy 1543 IREKALQOIQ-HLDYITQACLKNK-DVAIIELNENEE--NOVTRDIVENWSVEEQAME 1598
Db 1011 IRESOLKERLEKRYODIITHSLHARKPNLGLKSCDGNQELNFLRSLDIE--VATRLQK 1067
Qy 1599 MDLSVLOQVEDLERVAS-----ASLQ-----VKGWMCPEPASEREDLVYFE 1640
Db 1068 GGLGVETSEFEARVISLEKLDKFGECVIALQASVIKKFLOGFNAPKQ----- 1116
Qy 1641 HKSFTLCKEHDGFTGEDESSAHLERKSDNPLDIATRLADLERNIERIEEDIAPGL 1700
Db 1117 -----KRLQSEDSAKTEEVDEKQWVE--EAKVASAL 1148
Qy 1701 RVRRALSEARSAAQVALCICQLOKSIAWEKSIMKVYQICRKGDNELLLLCDCDKGC 1760
Db 1149 EKWKTAIREAQTFSRMHVLLGLMDACIKWDSAEARNCKVCPKKGEDDKLILDCENKAF 1208
Qy 1761 HTYCHRPITITPDGDFWPCACIACASGOTLIKIKLHVKGKKTNSKKGKVVLTGDTED 1820
Db 1209 HLCRLPALYEVFDGEWQCPACOPATA-----RRNSRGRNYTEESASESEDESDEE 1261
Qy 1821 EDS-----ASTSSSLKRGKNDLQKRKMEENT-SIN 1849
Db 1262 EEEEEEEEDYEVAGRLRPRKTIRGKHSHVIPPAAARGRRPGKKPHSTRSQKAPVD 1321
Qy 1850 LSKQESFTSVKPKRDDSKDALCSMILTEMETHEDANFPFLPVNLKLVPGYKVKIKPM 1909
Db 1322 AEVDLVLQTKRSSRQSLQKBEILHKIVKRYRFSWFPREPVRDEADYDYVITHPM 1381
Qy 1910 DFSTIREKLSSQYENLFEALDVLRLVFDNCETEN 1944
Db 1382 DFQTVQNKSCGSRVSVQFELTDMKQVFTNAEVYN 1416

RESULT 28

US-10-376-537-69
; Sequence 69, Application US/10376537
; Publication No. US20030224405A1
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042001
; CURRENT APPLICATION NUMBER: US/10/376,537
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US/09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 69
; LENGTH: 1525
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-376-537-69

Query Match 5.3%; Score 538.5; DB 15; Length 1525;

Best Local Similarity 18.1%; Pred. No. 1.2e-17;

Matches 353; Conservative 253; Mismatches 604; Indels 745; Gaps 64;

Qy 189 KPLSINVQAKETYMKLIVSPDVLKAGNKYTSSESSLLTSELRSKREY-----KQA 241

Db 8 KPFLVNLPGEEP-FTIPH-----TQEAFTREYEARLERYSERIWTCKST 55

Qy 242 FPSQKKQESSKSLKVIJAAALSNPKATSSSPA-HPKQTLNHNHPNPFLLTNALLGNHPNG 300
Db 56 GSSQLTHKEAEBEQEVAELLK-----BEFPAWYKLVLEMVHHN-----TA 97
Qy 301 VIOSVIOEAPLALTTKTMQSKINENIAAASSTPSSPVNLSTSGRRTPGNQTPVMPSPA 360
Db 98 SLEKLVDTAWLEIMTKYAVGECDFEV----- 124
Qy 361 PILHSQKKEKAVSNVNPVKTHSHHPAKSLVBOFRGTDSIPSSKDSNEDNEEDDE 420
Db 125 -----GKEKVL--KVKIVKI--HPL-----EKVDEE 146
Qy 421 EEDREDDDDSDSSESDESSEDTEGSEEDDDDDKQDESDD-----T 467
Db 147 ATEKKSAGCSPSDKENSQIAQDHQKQETVYKVEDGRRESINRARRSPRKLPTSUK 206
Qy 468 EGE-----KTSMKL--NKTSSVKSPPMSLTGHSTPRNLHAK----- 503
Db 207 KGERKWAAPPKPLPHKYDVKLQNEDEKISNV--PADSLIRTERPPNKEIVRIFIRNALRA 264
Qy 504 -----AFGSAPAALCSBSQSPA-----FLGTSSSTL--TSSPHSGTSKRRVVT 544
Db 265 GTGENAPWVVEDELVKYSLSKFSDFLLDPKYMTLNPSTKRKNTGSPDRKPSKSK-T 323
Qy 545 DERELRIPLEYGMQRETRIRNFGRLQGEVAYAPCGKKLQYPEVIKYLRSNGIMDISR 604
Db 324 DNSLSLSP-----NPKLWCHVHLKSL----- 347
Qy 605 DNFPSAKIRVGDVFEARDGPQEMQWCL-----LKEEDVIPIRAMEGRRGRPNDR 657
Db 348 -----GSPKLVKYNKSNKSPSEHEEEMKMSPNKLTNPHIPK-----KGP-PAK 392
Qy 658 QRAREESRMRRKGRPNVGNABFLDNADAKLLRLQAEIARQAAQIKLLRKLQKQQA 717
Db 393 KPGKSHDKPLKAKGRSGI-----LNGQKST 418
Qy 718 RVAKEAKKQQAIAAEBEKQKBOIKIMKQOEKIKRIOQ-----IRMEKEL--RAQ 766
Db 419 GNSKSPKK--GLKTPKTKMKQMTLLDMAKGTQKMTAPRNSGGTPTRTSSKPHKHLPPAAL 476
Qy 767 QILEAKKKKKEE-----AANAKLLEAEKRIK-EKEMR--RQAVLLKHQER-- 809
Db 477 HLIAYYKENDREDKRSALSCVISTARLLSSEDRARLPBELRSVLQRYELLEHKKWA 536
Qy 810 ---ERRRHQMLMKAMEARKKAEKERLKQEKREKLEQRRLLEMAKELKKP 866
Db 537 SMSEQRKEYLKKREBELKKLKEKAKERREKEMLERLEKQRYEDQEL---TGKVL--- 590
Qy 867 NEDMCLADQKPLPELPRIPLGLVLSGSTFSDCLMVVQFLRNFQKVLGFDVNDVNLVLQ 926
Db 591 -----PAPRLVDTPG-GLPNTLFGDVAMVVEFLSCYSGLLLPDAQYPITAVS-LM 638
Qy 927 EGLNIGDSMGEVQDILLVRLLSAAVCDPGLITGYKAKTALGHEHLLNVGNVNRDNVSEILQI 986
Db 639 EALADKGGFL-YLNRVLVILLQTLQDE-IAEDY---GELGMKLSIEPLTLHVSSELVRL 693
Qy 987 FM-----EAHCGQTELTESLTKAFOAHTPAOKASVLAFLIN 1023
Db 694 CLRRSDVQSESESGSDTDNDKDSAAFEDEVQDEFLKLETSEFFELTSEEKLIULTALCH 753
Qy 1024 ELACSKSVVSEIDKNIDYMSNLRDKWV-----EGKLRKLRIIHAKTGKRDTSGGIDLG 1079
Db 754 RILMTYSVQDHMETRQQAELWKEKRAEKLAVLKEENDKKRAEKQKEMAKKNGKENG 813
Qy 1080 EEQHPGLGTPGKRRKRGSDSDYDDDDDDSD-----QGEDDEDEEDEKE 1126
Db 814 -----LG-----KTRDKKRIYKPEQVDTAEADMI SAVKSRRLLAQAKKERIQE--R 860
Qy 1127 DQKGTDTICEDEDEGQQAASVELEKQIEKLSQOQSYRRKLFDAHSLSRSMVFGPDY 1186
Db 861 EMVKLERQAEERIRKHAAAEEK-----AFQEGIAKAKLVMRTIGTDNR 907
Qy 1187 RRRYWL-PRCGGIFVBMESGEGLEETAKEREKLKKAESVQIKKEEMFTSG--DSLNC 1242

```
Db 908 HNRVLPDSVDEGLFIE-----KGWVHDSIDY 934
QY 1243 SNTDCEQKEDLKEDONTNLFLOKPGSFKLSKLELVAKMPPESEVMTPKPNAGACGTL 1302
Db 935 RFNHH-----KDHT-----VSGCED 950
QY 1303 SYQNSGKSLGSVOSTATQSNVEKADSNLNTGSSGPGKFPYPLPNDQLLKLTEKNRQ 1362
Db 951 YCPRS KKAANLKNASMTQHG-----TATE----- 975
QY 1363 WFSLLPRTPCDDTLTHADMSTASLTVPQSPQSPKSPPTAPLGSQAQNPVGLNPFALS 1422
Db 976 ----- 975
QY 1423 PLQVKGVSMMGLQFCGWPTGVTSNPTFTLSVPSLSGLGSEGNNGSFLTNSVASSKS 1482
Db 976 ----- 975
QY 1483 ESPVPQNEKATSAQPAAVEVAKPVPDFSPKPIPEEMQFGWRIIDPEDLKALKVLHLRG 1542
Db 976 -----VAVETT-----TPKQGQNLWFLCDSQKELDELLCLHPQG 1010
QY 1543 IREKALQKIQK-HLDYITQACLKNK-DVAIIELNENEE--NQVTRDIVENWSVEEQAME 1598
Db 1011 IRESQLEKREKRYQDIHLSHLARKPNLGLKSCDGNQOELLNFLRSLDIE--VATRLQK 1067
QY 1599 MDLSVLOQVEDLERRVAS-----ASLQ-----VKGWMCPEPASERDLVYFE 1640
Db 1068 GGLGVVESTSEFARVISLEKLDKDFGECVIALQASVIKFLQGFMAPKQ----- 1116
QY 1641 HKSFTLCKEHGFTGEDESSAHLERKSDNPLDIAVTRLADLERNIERIEDIAPGL 1700
Db 1117 -----KRRKQSDSAKTEEVDEBEKKWVE--EAKVASAL 1148
QY 1701 RVWRALSEARSAAQVALCIIQOLQKSIWEKSIIMKVYCOICRKGDNBELILLCDGCKGC 1760
Db 1149 EKWTAIREAQTFSRMHVLLGMDACIKWDSAEANARCKVCPKKGEDDKLILDCENKAF 1208
QY 1761 HTYCHRPKITTPGDWFCPACIAKASQGTUKIKKLVHVGKKTNSKGGKVVTLTGDTED 1820
Db 1209 HLFCLRPALYEVDPGEWCQACAPATA-----RRNSRGRNYTEESASEDESEDEE 1261
QY 1821 EDS-----ASTSSLRKGNKDLQKRKWEENT-SIN 1849
Db 1262 EEEEEEEYEVAGLRLRPKRTIRGKHSVIPPAAASRRPGKPKHSTRRSQKAPPVD 1321
QY 1850 LSKQESFTSVKPKRDDS KDIALCSMILITEMETHEDAWPFLLPVNLKLVPYKVKVKKPM 1909
Db 1322 AEVDLVLQTKRSRRQSLQKCEELHKLIVKRYFSWPFPPEVTRDEAEDYDVITHPM 1381
QY 1910 DFSTIREKLSGGQVPLETTFALDVLRLVPDNCETFN 1944
Db 1382 DFQTVQNKCSGYSRVSQVEFLTDMQVFTNAEVN 1416
```

RESULT 29

```
US-10-702-148-68
; Sequence 68, Application US/10702148
; Publication No. US20040063145A1
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042002
; CURRENT APPLICATION NUMBER: US/10/702.148
; CURRENT FILING DATE: 2003-11-05
; PRIOR APPLICATION NUMBER: PRIOR APPLICATION NUMBER: US/09/839,479
; PRIOR FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JF98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
```

```
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 68
; LENGTH: 1525
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-702-148-68
```

```
Query Match 5.3%; Score 538.5; DB 15; Length 1525;
Best Local Similarity 18.1%; Pred. No. 1.2e-17;
Matches 353; Conservative 253; Mismatches 604; Indels 745; Gaps 64;
```

```
QY 189 KPSLVNQAKKETYMKLIVFSPDVLKAGNKNTSESSLTSELKSKREQY-----KQA 241
Db 8 KPFLVNLPLGEEPP-FTIPH-----TOEAFRTREYEYARLERYSERIWTCKST 55
QY 242 FPSOLKKEQSSKSLKKVIAALSNNPKATSSSPA-HPKQOTLENNHNPFLTNALLGNHQPNG 300
Db 56 GSSQUTHKAEWEEQEAELK-----EFPAMYTEKLVLEMVHN-----TA 97
QY 301 VIQSVIOEAPLALTTTKMQSKINENIAAASSTPFPSPVNLSTSGRRTPGNGQTPMPSAS 360
Db 98 SLEKLVDTAWLEIMTKYAVGECDFEV----- 124
QY 361 PILHSQKEKAVSNVNPVTOHHSHPAKSLVEQFRGTDSDIPSSKDSSEDSNEDEEDE 420
Db 125 -----GKERML--KVKIVKI--HPL-----EKVDEE 146
QY 421 EDEDEDEDESDSQSESNSSESDETEGSEEDDDDKQDESDD-----T 467
Db 147 ATEKSDGACDSPSDKENSQIAQHQKKEVTVKEDEGRRESINDARRSPRKLPTSLK 206
QY 468 EGE-----KTSMKLJ---NKTTSVVKSPMSLTGHSTPRNLHIAK----- 503
Db 207 KGERKWAPPKFLPHKYDVKLQNEDKIISNV--PADSLIRTERPPNKEIVRYFIRHNALRA 264
QY 504 -----APGSPAALCSESQSPA-----FLGTSSTL-----TSPSHGTSKRRVVT 544
Db 265 GTGENAIVWVEDELVKKYSLSKFSFDFLDPYKTYMTLNPSTKRNKNTGSPDKPKSKSK-T 323
QY 545 DERELRIPLEVGMQRETRIENFGGRLQGEVAYAPCGKLLRQYPEVILKYSRNGIMDISR 604
Db 324 DNSSLSPL-----NPKLWCHVHLKKSLS----- 347
QY 605 DNFSFAKIRVGDYPFYEARDGPQEMQWCL-----LKEEDVIPRAMEGRGRPPNPDR 657
Db 348 -----GSPLVKVNSKNSKSPPEHLEEMMKWSPNKLHTNPHIPK-----KGP-PAK 392
QY 658 QRAKEBSRMRRRKGRPPNVGNVNAEFLDNADAKLLRKLQAEIARQAAQIKLLRKLQKQEQ 717
Db 393 KPGKHSKPLKAKGRSGI-----LNQKST 418
QY 718 RVAKEAKQQAIAAAEKKQKEQIKIMQOKEIKRIQ-----IRMEKEL--RAQ 766
Db 419 GNSKSPKK--GLTKPTKMKQMILLDMAKGTQKMTAPRNSGGTPRTSSKPKHKLPPAAL 476
QY 767 QILEAKKKKKEE-----AANAKLLEAEKRIK-EKEMR---RQOAVLLKHQER-- 809
Db 477 HLTAIYKENKDRKRSALSCVISKTARLLSSSEDRARLPEELRSLVQKRYELLEHKKWA 536
QY 810 ---ERRQHMMLMKAMEARKKAEKERLQKQKDEKRLNKERKLEQORRLEUEMAKELKKP 866
Db 537 SMSBEQKRYLKKRBEELKKLKEKAKERKEKEMLERLEKQRYEDQEL---TGKNL--- 590
QY 867 NEDMCLADQKPLPELPRIPGLVLSGSTFSDCLMVVQFLRNFQKVLGVDVNDVNLVLQ 926
Db 591 -----PAFRLVDTPPE-GLPNTLFGDVAMVVEFLSCYSGLLLPDAQPITAVS-LM 638
QY 927 EGLNIGDSMGEVDLLVRLLSAAVCDPGLITGYKAKTALGELHLLNVGNRDNVSEILQI 986
```

```
Db 639 EALADKGGFL-YLNRVLVILLQTLQDE--IAEDY----GELGMKLSEIPIILTHSVSELVRL 693
QY 987 FM-----BAHCGQTETESLTKAFOAHTPAQKASVLAFILN 1023
Db 694 CLRRSDVQSESGSDTDDNKDAAPEDEVQDEFLEKLETSEFFELTSEKQLILTALCH 753
QY 1024 ELACSKSVSEIDKNDIDYNSLNRDQWVY----EGKLRKLRILHAKTKGRDTSGLDGLG 1079
Db 754 RILMTYSVDHMETRQMSAELWKERLAVLKEENDKKRAEKQKREMEAKNKENGKVGNG 813
QY 1080 BEQHPLGTPTPKRRKRRKGGSDYDDDDDDDDDD-----QGDDEDEDEDEKE 1126
Db 814 -----LG-----KTRKKRIVKFFPOVDTAEADMISAVKSRRLALIAQAKEREIQE--R 860
QY 1127 DQKGGKTDICEDDEGDQAAASVEELEKQIEKLSKQOSQYRRKLFDAHSLSRSMFGPDQRY 1186
Db 861 EMKVLERQAEERIKKHAAEK-----AFQEGIAKAKLVMRRTPIGTDRN 907
QY 1187 RRRYML-PRCGGIVFEGHESGEGLEEIAKEREKUKAESVQIKSEMPFTSG---DSLNC 1242
Db 908 HNRVWLFSDVPGLEFIE-----KGMVHDSIDY 934
QY 1243 SNTDCEQKEDLKEKONTNLFLOKPGSFSKLSKLLLEVAKWPESEVMTPKPNAGANGCTL 1302
Db 935 RFNHC-----KHHT-----VSGCED 950
QY 1303 SYQNSGKHSGLSVQSTATQSNVEKADNNLNTGSSGPKFYSPLPNDQLLTKLTEKNRQ 1362
Db 951 YCPRSKKANLGNASNTQHG-----TATE----- 975
QY 1363 WFSLLPRTPCDDTSLTHADMSTASLVTPOSQPSKSPSPTRAPLSSAQNVPGLNPFALS 1422
Db 976 ----- 975
QY 1423 PLQVKGVSMMGLQFCGMPGVVTSNIPFTLSVPSLGSGLGSEGNNSFLTNSVASSKS 1482
Db 976 ----- 975
QY 1483 ESPVPQNEKATSAQAAVEVAKVPDPSPKPIPEEMQFGWRIIDPEDIKALLKVLHLRG 1542
Db 976 -----VAVETT-----TPKQGNLWFLCDSQKELDELLNCLHPQG 1010
QY 1543 TREKALQKQIQ-KHLDYTOACLKWK-DVAIIELENEE--NOVTRDIVENWVSBEQAME 1598
Db 1011 IRESOLKERLEKRYQDIIHSLHARKPNLGLKSCDGNQELLNFLRSLDIE---VATRLQK 1067
QY 1599 MDLSVLQVQEDLERVAS-----ASLQ-----VKGMCPPEPASEREDLVYFE 1640
Db 1068 GGLGVSETSFEFARVILEKLDKDFGECVIALQASVIKFLQGFMAPKQ----- 1116
QY 1641 HKSFTKLCKEHGDEFTGEDESSAHALERKSDNPLDIAVTRLADLERNTERRIEEDIAQGL 1700
Db 1117 -----KRLQSEDSAKTEEVDEKKWVE--EAKVASAL 1148
QY 1701 RVWRALSEARSAAQVALCIOOLQKSIaweKIMVYQICRKGDNBELLLLCDGCKGC 1760
Db 1149 EKWKTAIREAQTFRMHLVLLGMDACIKWDSMAENARCKVCPKGGEDDKLILCDCECNKAF 1208
QY 1761 HTYCHRPKITTPDGDWFCPACIAKASGOTLKIKLHVKGKKTNSKGGKVVTLTGDTED 1820
Db 1209 HLFCLRLPALYEPDGEWQCPACQAPATA-----RRNSRGRNYTSESASEDESDSEBEE 1261
QY 1821 EDS-----ASTSSSLKRGKNDLQKRKWEENT-SIN 1849
Db 1262 ESEEBEEDYEYVAGRLRPRKTIIRGKHSVIPPARGSGRRPGKKPHSTRSQPKAPPVD 1321
QY 1850 LSKQESFTSVKPKRDDSODLALCSMILTEMETHEDAWPFLLPVNLKLVPGYKVKIKKPM 1909
Db 1322 AEVDLVLQTKSSRRQSLELQKCEILLHKIVKXRFSPFPFREPVTDEAEDYVDVITHPM 1381
QY 1910 DFSTIREKLSGGQYNLETFALDVLVFDNCFTFN 1944
Db 1382 DFQTVQNCSCGYSYRQVEFLTDMQVETNAEVYN 1416
```

RESULT 30

```
US-09-839-479-29
; Sequence 29, Application US/09839479
; Publication No. US20020039779A1
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042002
; CURRENT APPLICATION NUMBER: US/09/839,479
; CURRENT FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 29
; LENGTH: 1531
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-839-479-29
```

Query Match 5.3%; Score 533; DB 9; Length 1531;

Best Local Similarity 17.9%; Pred. No. 2.2e-17;

Matches 352; Conservative 256; Mismatches 600; Indels 756; Gaps 67;

```
QY 189 KPLSLVQAKKETYMKLIVPSPDVLKAGNKTSESSLLTSELRSKREY-----KQA 241
Db 8 KPFPLVNPLOEPP-FTIPH-----TQEAFTREYEARELERSYRITWCKST 55
QY 242 PFSQLKKQESSKSLKVIAALSNPKATSSSPA-HPKQTLNNHPNPLTNALLGNHPNG 300
Db 56 GSSQLTHKEAWEEQEVAELEK-----EFPAYEKLVLVEMVHN-----TA 97
QY 301 VIQSVIQAPLALTTKTKQSKINENIAAASPTSPSSPNLSTSGRTTGNQTPVWPSAS 360
Db 98 SLEKLDVDFAWLEIMTKYAVGECDFEV----- 124
QY 361 PILHSQGEKAVSNVNPVKTHSHPAKSLVEQPGRTSDIPSSKDSDESDNEDEDE 420
Db 125 -----GKEML--KVKLVKI--HPL-----EKVDEE 146
QY 421 EDEEDDDDESDSDSESDESDTEGSEEDDDDDKQDESDSD-----T 467
Db 147 ATEKSDGACDPSDDKENSQIAQDHQKKTVVKEDSGRRRESINDRARRSPKRLPTSILK 206
QY 468 EGE-----KTSMKL---NKTTSVKSPSMSTLGHSTPRLHIAK----- 503
Db 207 KGERKWAAPPKFLPHKYDVKLQNEDKIISNV--PADSLIRTERPPNKEIVRVFIRNALRA 264
QY 504 -----APGSAAPALCSESQSPA-----FLGTSSSTL---TSSPHSGTSKRRRV 544
Db 265 GTGENAPVWVDELVKYSLSKFSDFLLDPYKMTLNPSTKRNKVTGSPDRKPKSKKST 323
QY 545 DERELRIPLEYG-WORETRIRNFQG---RLQGEVAVYAPCGKKLQRYDEVIKYLSRNGIM 600
Db 324 DNSSLSPFLNPKLWCHVHLKKSLSGSLPKVKNKSKSP-----EBHLEMMKQWSPN--- 376
QY 601 DISRDNFSFAKIRVGDFFYEARDGQEQEMQWCLLKEEDVPIPRAMEGRGRPPNDRORA 660
Db 377 -----KLHT-NPHIPKKG- 395
QY 661 REESMRRRKGRPPNVGNAGAEFLDNADAKLRLKLAQAEIARQAQIKLLKLOKQSOARVA 720
Db 396 KHSKPLKAKGRSKGI-----LNGQSTGNS 421
QY 721 KEAKKQQAIAAAEEKKQKEQIKIMQOEKIRIQO-----IRMEKEL--RAQQL 769
```

Db	422	KSPKK--GLKTPKTKMQMTLLDWAQTKQKTRAPRNSGGTPTSSPXPKHLPPAALHLI	479
Qy	770	EAKKKKEE-----AANAKLAEAKRIK-EKEMR---RQAVLLKHOER-----	809
Db	480	AYYKENDREDKRSALSVCVISKTRLJSSDRAKLPEELASLQVKRYELHEHKRWASMS	539
Qy	810	ERRQHMMLKAMEARKABEKERLQKQEKDEKRLNKERKLEQRRLEMAKELKKPNE	869
Db	540	EEQRKEYLKKREBLKKLKEKAKERREKEMLERLEKQRYEOEL---TGKNL-----	590
Qy	870	MCLADQKPLPELPRIPLGLVJSGSTFSOCLMVQVPLRNFVKVLGPDVNNIDVNLVQEG	929
Db	591	-----PAPRLVDTPG-GLPNTLFGVAMVVEFLSCYSGLLLPDAQYPIITAVSLMEALS	642
Qy	930	LNIGD--SMGEVQDILLVRLLSAAVCDPGLITGVYKAKTALGEBHLLNVGNRDNVSEILQIF	987
Db	643	ADKGGFLYLNARVLVILLQTLQTLQDEIADY---GELGNKJSEIPLTHUSVSELVRLC	699
Qy	988	M-----EAHCGQTELTESLTKTFAQHAHTPAQKASVLAFLINE	1024
Db	700	LRRSDVQEESEGSTDDNKDSAAPEDEVQDEFLKLETSFEFFELTSEEKLIQILTALCHR	759
Qy	1025	LACSKSVVSEIDKNIDVMSLNRDKWV---EQRKRLRIIHAKTGKRDTGGIDLGE	1080
Db	760	ILMTYSVQDHMETRQQAELWKERLAVLKEENDKRAEKQKREMAKNKENGKVGENG-818	
Qy	1081	EQHPLGTPPCRERRKGGSDYDDDDDDSDD-----QGDEDEDEEDKED	1127
Db	819	-----LG-----KTRDKRIRVKFEPQVDTEAADMISAVKSRRLLAIOAKKERIQE--RE	866
Qy	1128	QKQKKTIDICEDEBQGAASVEELEKQIEKLSQSOYRRKLPDASHLSRSMVFGPDYR	1187
Db	867	MKVLELQAEERIRKHKAAEK-----AFQGIKAKALVMRTTIGITDRNH	913
Qy	1188	RRYWL-PRCGIFVEGMESEGLEETAKEREKLIKAEVSQIRKEEMPETSG---DSLNC	1243
Db	914	NRWLFSDVEPGLFIE-----VSGDEDYCPRSK-----KANLGKNA-----	969
Qy	1244	NTDCEQEKEDLKEKDNLTNLFLOKPGSFSLKSLLEVAKMPPESEVMTPKPNAGANGTLS	1303
Db	941	FNHHC-----KDHT---VSGDEDYCPRSK-----KANLGKNA-----	969
Qy	1304	YQSGKHSGLSGVQSTATQSNVERADSNLNTGSSGPGKFVSLPNDQLLTKLTKEURQ-	1362
Db	970	-----SMYTOHGTATEVAE-----TTTPKQGN	993
Qy	1363	WFSLLPRTPCDDTSLTHADMTASLVTPOSPPSKSPSTPAPLGSQAQNPVGLNPPAL	1421
Db	994	LWFL-----CD-----	999
Qy	1422	SPLOQKGVMMGLQFCGWTGVVTSNIPFVLSVPSLGSGLGISEGNGSNFSLTSNVASSK	1481
Db	1000	-----	999
Qy	1482	SESPVQNEKATSAQPAAVEVAKVPDPPSPKPIPEEQFGWRIIDPEDLKALLKVLHLR	1541
Db	1000	-----SOKLEDELNCLHPQ	1014
Qy	1542	GIREKALQKIQK-HLDYITQACLKNK-DVAIIELNENE--NQVTRDIVENMSVEEQAM	1597
Db	1015	GIRESQLKERLEKRYQDIHSHILARKPNLGLKSCDGNQELLNFLRSDLIE---VATRLQ	1071
Qy	1598	EMDLSVLQVEDLERRVAS-----ASLQ-----VKGWMCPEPASEREDLVYF	1639
Db	1072	KGGLGYVEETSEFAFRVISLEKLDKDFECVIALQASVIKFFLOQFMAPQ-----	1121
Qy	1640	EHSFTKLCXHDGEFTGEDESSAHALERKSDNPLDIATVRLADLERNIERREEDIAPG	1699
Db	1122	-----KRRKJSEDSAKTEFVDEEKKWVE---EAKVASA	1152
Qy	1700	LRVRRRLSEARSAAQVALCIIQQLQKSIKAEKSIKMYQICIRKGDNEELLLLDCGCDKG	1759

RESULT 31

US-10-376-537-29

Sequence 29, Application US/10376537

; Publication No. US20030224405A1

; GENERAL INFORMATION:

APPLICANT: Jones, Michael H.

: TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR

FILE REFERENCE: 06501-042001

FILE REFERENCE: 06301-042001
CURRENT APPLICATION NUMBER: US/10/376.537

;; CURRENT APPLICATION NUMBER: US/10/376,333
: CURRENT FILING DATE: 2003-02-28

; CURRENT FILING DATE: 2003-02-28
 ; PRIOR APPLICATION NUMBER: IIS/09/418-710

; PRIOR APPLICATION NUMBER: US/0
: PRIOR FILING DATE: 1999-10-15; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01782

PRIOR APPLICATION NUMBER: PCI/17-1988-04-17

; PRIOR FILING DATE: 1998-04-17
 ; PRIOR APPLICATION NUMBER: 1998/010027

; PRIOR APPLICATION NUMBER: JP 9/310027
 ;

; PRIOR FILING DATE: 1997-10-24

; PRIOR APPLICATION NUMBER: JP

; PRIOR FILING DATE: 1997-

; NUMBER OF SEQ ID NOS: 73

; SOFTWARE: Fas

; SEQ ID NO 29

; LENGTH: 1531

; TYPE: PRT

; ORGANISM: HO

Query Match 5.3%; Score 533; DB 15; Length 1531;

Query Match 3.3%, Score 335, DB 13
Best Local Similarity 17.9%: pred. No. 2.2e-17:

Best Local Similarity	17.9%	Fied: NO.	2.2e-17,		
Matches	352:	Conservative	256: Mismatches	600:	
		Indels	756:	Gaps	67:

189 KPLSLVNOAKKETYMKLIVPSPDVLKAGNKNTSESSLTSELRSKREQ-----KQA 241

[illegible]

242 FPSOI.KKOESSKSI.KKVTAAL.SNPKATSSSPA-HPKOTI.ENNTHPNPEI.TNAI.I.GNHOPNG 300

QY 242 FFSQLKRRQJESSKSLKKVIAALSNPRAITSSSPA-HPRQVILENNHHPNPFELINALLGNHQPNQ 300

Q-: 301 VITACONCEAD? A? I? TTTTWTWMOCKYNTNENT A A SCOTEECCDIANT CTSOCBPBTDNCNOTBIWMDLRS 760

301 VIQSVIQEAPLALTTTKIKMQSKINENIAAASSTFFSSPVNLSTSGRRTPGNQTFFVMPAS 360 QY

C

QY 361 PILHSQGEKAVSNVNPVKTQHSHHPAKSLVEQFRGTDSDIPSSKDSNDEDEEDDE 420

QY 421 EEDEDEDEDDSQSESDNSSESDTEGSEEEEDDDKQDESDSD-----T 467

Qy 468 EGE-----KTSML---NKTTSVKSPMSLTGHSTPRNLHIAK----- 503

QY 504 -----APGSAPALCESOSPA-----FLGTSSSTL-----TSSPHSGTSKRRVVT 544
Db 265 GTGENAPWVVEDELVKKTSLSKFSDFLLDPYKNTLNPSTKRNKTGSPDRKPSKSK-T 323
QY 545 DERELRIPLEYG-WQRETRIRNFGG---RLQGEVAYAPCGKLLQYDEVIKYLSENGIM 600
Db 324 DNSSLSPLNPKLWCHVHLKKSLSGSPUKVKNKSNKSGP---EHLSEMMKMSFN--- 376
QY 601 DISRDNFSFSAKIRYGDVEARDGQPMQWCLLKEEDVIPRIRAMEGRGRPPNPDRORA 660
Db 377 -----KLHT-NFHI PKKGP-----PAKPG 395
QY 661 REESMRARRKRPNNVGNABFLDNADAKLLRLKLAQETARQAQAQLLRKLOKQOARVA 720
Db 396 KHSDXPLKAKGRSGI-----LNGQKSTGNS 421
QY 721 KEAKQQAIAAAEERKKEQKIKMKQOEKIKRIOQ-----IRMEKEL--RAQOIL 769
Db 422 KSPKK--GLKTPKTKMKTLLDMAKGTQKMTAPRNSGGTPRTSSKPHKLLPPNALHLI 479
QY 770 EAKKKKEE-----AANAKLLEAEKRIK-EKEMR---RQOAVLLKHOER----- 809
Db 480 AYYKENKREDKRSALSVCISKTARLSSSEDRAFLPEELRSILVQRYELLEHKKEWASMS 539
QY 810 ERRORHMLMKAMEARKKAEBERUKQEKDEKRLNKERKLEQORLELEMAKELKKPNED 869
Db 540 EQRKEYLLKKEELKKLKEKAKERRERKEMLERLEKQKVEDQEL---TGKNL----- 590
QY 870 MCLADQKPLPELPRIPGLVLSGSTSDCLMVVOFLRNFVKVLGFDVNDVNLVLQSLG 929
Db 591 -----FAFRVLVDTFE--GLPNTLFGDVAMVVEFLSCYSGLLPDAQYPIITAVSLMEALS 642
QY 930 LNIIGD--SWGEVQDILLRLLSAACDPGLITGYKAKTALGEHLLNVGNRDNVSILOIF 987
Db 643 ADKGGFLYNRLVLILLQTLQLQDEIAEDY---GELGMKLSIPLILHSVSELVRLC 699
QY 988 M-----BAHCGQTELTSKTKPAQAHTPAKQASVLAFLINE 1024
Db 700 LRRSDVQBESEGSTDDNKDSAAFEDNEVDQEFLEKLETSFEFFELTSEBKLIQTALCHR 759
QY 1025 LACSKSVUSEIDKNIDYNSLRDRKVV-----EGKLRKRIIHAKTGKRTSGSIDGE 1080
Db 760 ILMTYSVDQMETROQMSAELWKERLAVLKEENDKKRAEKQKREMEAKNKGKVENG- 818
QY 1081 EQHPLGTTPGKRRRRKGDSDYDDDDDDSD-----QDEDEDEDEDED 1127
Db 819 ----LG-----KTRKKRIYKFEQVQVTEAEDMSIAVKSRRLLAIQAKKEEIQE--RE 866
QY 1128 QKGKTDICEDDEGDAASVEELEKQIEKLSKOOSQYRRKLFDAHSLRSVMFGPDYR 1187
Db 867 MKVKLERQAEERIRKHAEEK-----AFQEGIAKAKLVMRRTPIGTDRNH 913
QY 1188 RRYWIL-PRCGGI FVEGHESGEGLEIAEKEREKKAESVQIKKEEMPETSG---DSLNC 1243
Db 914 NRYWLFDEVPLGTFIE-----KGWVHDSIDYR 940
QY 1244 NTDHCQEKEDLEKONTNLFLOKPGSFSLKLELVAKMPPESVMTPKNAGANGCTLS 1303
Db 941 FNHHC-----KDHT---VSGDEDYCPRSK-----KANLGKVA----- 969
QY 1304 YONGSKHLSGVQSTATOSNVEKADSNLNFNTSSGGPGKFYSPLPNDQLLTKLTKNRQ- 1362
Db 970 -----SNWTOHGATAVE-----TTTPKQGN 993
QY 1363 -WFSLLPRTPCDDTSLTHADMSTASLVTPOSPQSPSPPTAPLGSSAQPNVGLNPPAL 1421
Db 994 LMFL-----CD----- 999
QY 1422 SPLQYKGVGMGLQFCGWPTGVTSNIPFTLSVPSLGLSLSGNGNSFLTSTNVASSK 1481
Db 1000 ----- 999
QY 1482 SESVPQNEKATSAQAAVEVAKPVDPPSPKPIPEMQFGWRRIIDPDKALLKVLHLR 1541

Db 1000 -----SOKELDELNCLHPQ 1014
QY 1542 GIREKALQKIQK-HLDVITQACLKKNK-DVAIIENENEE--NOVTRDIVENWSVEEOM 1597
Db 1015 GIRESQLKERLEKRYQDIIHSIHLARKPNLGLKSGDGNQELLNPLRSLEIE---VATRLQ 1071
QY 1598 EMDLSVLQOVBDLERRVAS-----ASLQ-----VKGMWCEPASEREDLYVF 1639
Db 1072 KGLGYVEETSEFEARVISLEKLDKDFGBCVIALQASVIKKFLOGFMAKQ----- 1121
QY 1640 EHKSFTKCKEHDGEFTGEDESSAHALERKSDNPLDIAVTRLADLERNIERIEBEDIAAPG 1699
Db 1122 -----KRRKLOSEDSAKTEEVDEEKKQVE--EAKVASA 1152
QY 1700 LRVMRRALSEARSAQAVALCIQLOKSTAWSEKSTMKVYCOICRKGDNBEELLLLCDGCDKG 1759
Db 1153 LEKWKTAIREAQTSRHHVLLGMLDACTKWDNSAENARCKVCPKKGEDDKLILCDECNKA 1212
QY 1760 CHTYCHRPKITIPDGWDFCPACIAKASGOTLKIKLHVKKKTNESKKGKVVTLTGDT 1819
Db 1213 FHLFLPALYEVDPGEWQCPACQPATA-----RRNSRGRNVTESASE-----DSE 1259
QY 1820 DEDS-----ASTSSSLKRGNKDLOKRM 1842
Db 1260 DDESEDEEEEBEEDYEVAGLRPRKTIKSHSVIPPAARSGRRPGKKPHSTRS 1319
QY 1843 EENTS--INLSKQSFSTSVKKPKRDDSKDALCSMLTTEMETHEDAWPFLPVNLKLVPG 1900
Db 1320 QPKAPPVDDAEVDELVLTQKSSRRQSLQKCEIHLKIVKYRFSWFPFVTRDEAD 1379
QY 1901 YKVIKPMDFSTIREKLSQOYPNLETFAIDVRLVDFDNCETFN 1944
Db 1380 YYDVI THMDPQTQVQNKCSGYSYRQVEFLTDMKQVFTNAEYVN 1423

RESULT 32

US-10-702-148-29
; Sequence 29, Application US/10702148
; Publication No. US20040063145A1
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042002
; CURRENT APPLICATION NUMBER: US/10/702,148
; PRIOR FILING DATE: 2003-11-05
; PRIOR APPLICATION NUMBER: PRIOR APPLICATION NUMBER: US/09/839,479
; PRIOR FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 29
; LENGTH: 1531
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-702-148-29

Query Match 5.3%; Score 533; DB 15; Length 1531;

Best Local Similarity 17.9%; Pred. No. 2.2e-17;

Matches 352; Conservative 256; Mismatches 600; Indels 756; Gaps 67;

QY 189 KPLSIVNQAQKETYMKLIVRSPDVLKAGNKVTSBESSLLTSELRSKREOY-----KOA 241

Db 8 KPFLVNPPLGEEPP-FTIPH-----TOEAFRTREYERLARLESRIWTCKST 55

QY 242 FPSQLKKQESSKLLKVIAALSNPKATSSSPA-HPKQTLNHNHPNFTLNALLGNHPNG 300


```
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 27
; LENGTH: 1527
; TYPE: PRN
; ORGANISM: Homo sapiens
US-09-839-479-27

Query Match
Best Local Similarity 18.0%; Score 532; DB 9; Length 1527;
Matches 354; Conservative 255; Mismatches 597; Indels 756; Gaps 68;

QY 189 KPLSLVNOAKETYMKLVIPSPDVLKAGNKTSSSESSLLTSELRSKREYQ-----KQA 241
DB 8 KFFLVNPLGPEPF-FTIPH-----TQFAFRTEYEARELERYSERIWTCKST 55
QY 242 PPSQLKQESSKSLKVAALSNPKATSSSPA-HPKQTLNNHPNPFLLTNALLGNHPNG 300
DB 56 GSSQLTHKEAWEEBEVALLK-----BEFPANYEKLVLVWVHN-----TA 97
QY 301 VIQSIOEAPLALTTTKRMQSKINENIAAASSTPFPSPVNLSTSGRTPGNQTPVMPAS 360
DB 98 SLEKLVDTAWLEIMTKYAVGECDFEV-----124
QY 361 PILHSGQKEKAVNNVNPVKTOHSHHPAKSLVQPRGTSDIPSSKSDSEDSDEEDDE 420
DB 125 -----GKEKML--KVKIYKI--HPL-----EKVDEE 146
QY 421 BEDEBDEDDSDSOSDSNSDSTEGSBEEDDDKDDSDSD-----T 467
DB 147 ATEKKSOGACDPSDDKSNSSQIAOHOKKTVVKEDEGRRESINDRARRPRKLPSTSLK 206
QY 468 EGE-----KTSMKL-----NKTSSVKSPMSLTGHSPTPRNLHIAK-----503
DB 207 KGERKWAPPKPLPHKYVDVKLQNEKIIINNV--PADSLIRTPPNKEIVRYFIRHNALRA 264
QY 504 -----APGSAPALCSQSQA-----FLGTSSSTL-----TSSPHGTSKRRVVT 544
DB 265 GTGENAPWVVEDELKYSKSLPSKFSDFLLDPYKMTLNPSTPKRKNVTGSPDRKPSKSK-T 323
QY 545 DERELRIPLYG-WQRETRIRNFGG---RLQGEVAYAPCGKLRQYEVLYKLSRNGIM 600
DB 324 DNSSLSSPLNPKLWCHVKHKKSLSGSLPKVKNKSKSP-----EBHLEEMKQMSPN---376
QY 601 DISRDNFSFSAKIRVDFYEARDGQEQWCLLKEEDVIPRAMEGRGRPPNPDORA 660
DB 377 -----KLHT-NFHI PKGP-----PAKPG 395
QY 661 REESRRRRKGRPPNVGNABFLDNADAKLLRLQAEIARQAAQIKLLRKLOKQOARVA 720
DB 396 KHSDKPLAKGRSGKI-----LNGQKSTGNS 421
QY 721 KEAKQQAIAAEERKQKEQIKIMQOEKIKRIQO-----IRMEKEL--PAQIIL 769
DB 422 KSPKK--GLTKPTKTKQKMTLLDMAKGTQKMTRAPNSGGTPTSSKPKHKLPPAALHLI 479
QY 770 EAKKKKKEE-----AANAKLLAEKRIK-EKEMR---ROQAVLLKHOER-----809
DB 480 AYYKENKREDRKSALSCVSIKTARLLSSEDPARLPEELRSVQRYELLEHKKQWASMS 539
QY 810 ERRROHMLMKAMEARKKAEKERLQEKREKRLNKERKLEQORLELEMAKELKKNED 869
DB 540 EBQREYKLGKREELKCKLKEKAEKREKEMLEKQKRYBDQEL---TGKVL-----590
QY 870 MCLADQKPLPELPRIPGLVLSGSTSDCLMVVQFLNFKGVLFQFDVNDVFNLSVLQGL 929
DB 591 -----PAFRLVDTPE-GLPNTLFGDVAWVVEFLSCYSGLLPDAQYPIITAVSLMEALS 642
QY 930 LNIQSGMEVQDVLRLLSAAVCDPLGITGYKAKTALGELLNVLGNVNDVNSEIILQIFM- 988
DB 643 ADKGGFL-VLNVVLVILLQTLQDE-IAEDY---GELGKLSSEIPLTHSVSELVRLCLR 697
QY 989 -----EAHCGQTELTESLKTKFAQHTPAQKASVLAFLINELA 1026
```

RESULT 34

```
US-10-376-537-27
; Sequence 27, Application US/10376537
; Publication No. US20030224405A1
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042001
; CURRENT APPLICATION NUMBER: US/10/376,537
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US/09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 27
; LENGTH: 1527
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-376-537-27

Query Match      5.2%; Score 532; DB 15; Length 1527;
Best Local Similarity 18.0%; Pred. No. 2,4e-17;
Matches 354; Conservative 255; Mismatches 597; Indels 756; Gaps 68;

QY 189 KPLSLVNAQAKETMYKMLVPSDVLKAGNKTSESSLLTSELRSKREY-----KQA 241
DB 8 KPFPLVNLPGSEPF-FTIPH-----TQEAFTREEYEARELRYSERIWTCKST 55
QY 242 PPSQLKQESSKSLKVVTAALSNPKATSSSPA-HPKQYLENNHPNPFITNALLGNHQPNG 300
DB 56 GSSQLTHKEAWEEQEVALLK-----BEFFAWYKLVLEMVHNN-----TA 97
QY 301 VIQSVIQAPALATTKMQSKINENIAAASSTPSSPVLNSTSGRTPGNQTWPMPAS 360
DB 98 SLEKVDVTAWEIMTYAVGECDFEV-----124
QY 361 PILHSQKKEKAVNNVNPVKTQHSHHPAKSLVEQPRGTDSIPSSKDSNEDEEEDDE 420
DB 125 -----GKEKML--KVKIVKI-----HPL-----EKVDEE 146
QY 421 EDEDEDDDDSDSQSSEDSNSSESDTEGSSEEDDDDDKQDESDD-----T 467
DB 147 ATEKSDGACDPSPSDKSNSQIAQHQKQETVVKEDGRRESINDRARRSPKPLPTSLK 206
QY 468 EGE-----KTSMKL--NKTSSVKSPMSLGHSTPRNLHIK-----503
DB 207 KGERKWAPPKPLPHKYDKVLQNEDKIISNV--PADSLRTRPPNKEIVRYFIRHNALRA 264
QY 504 -----APGSAPAALCSQSPA-----FLGTSSSTL--TSSPHSGTSKRRRTV 544
DB 265 GTGENAPWVEDELVKYSLSPKSFDFLLDPKYMTLNPSTKXKNTGSPDRKPSKSK-T 323
QY 545 DERELRIPLEYG-WORETRIRNFGG---RLOGEVAIYAPCKUKRQYDEVIKYLSRNGIM 600
DB 324 DNSSLSSPLNPKLWCHVLKXKLSLSGLSPKLVKNSKNSKSP---BEHLEENMKMSPN---376
QY 601 DISRDNFSFAKIRVGDPEYARDGQEMQWCLLKEEDVI PRIRAMEGRGRPPNDRORA 660
DB 377 -----KLHT-NFHI PKKGP-----PAKPG 395
QY 661 REESMRRRKGRPPNVGNAEFLDNADAKLLRLKLOAEIARQAAQIKLLRLKLOEQEARVA 720
DB 396 KHSKPLKAGRSKI-----LNGQKSTGNS 421
QY 721 KEAKKQQAIAAEERKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQ 769
DB 422 KSPKK--GLKTPKTKWQMTLLDMAKGTQKQWTRAPRNSGGTPTRTSSKPKHKLPPAALHLI 479
```

```
QY 770 EAKKKKEE-----AANKLLEAEKRIK-EKEMR---RQQAALLKHQER-----809
DB 480 AYYKXKDRKRSALSCVISTARTALLSSEDRARLPEELRSIVQKRELEHKKRWASMS 539
QY 810 ERRQHMMLKAMBARKKAEEKERLKOBRDEKINKERKLEQRRLEUEMAKELKPNED 869
DB 540 EQRKEVLLKKREELKKLKEKAKERREKEMLEKQRYEDQEL---TGKNL-----590
QY 870 MCLADQKPLPELPRI PGLVLTSGSTFSDCLMVUVQIRNFGKVLGPDVNDIVNLSVLQBL 929
DB 591 -----PAFRLVDTPF--GLENTLFGDVAMVVEFLSCYSGLLLPDAQYPIAVSLMEALS 642
QY 930 LNIQDSMEVQDVLRLLSAAVCDPGLITGYKAKTALGALHLLNVGNRDNVSEILQIFM- 988
DB 643 ADKGGLF--YLNRLVILLOTLLODE-TAEDY---GELGMKLSIEPLTLHSVSELVRLCLR 697
QY 989 -----EAHCGQTELESKTAKFAQHTPAQKASVLAFLINELA 1026
DB 698 RSDVQEESESGSDTDNDKDSAAAFEDNEVQDEPLEKLETSEFFELTSEKQLTALTALCHRL 757
QY 1027 CSKSVVSEIDKNIDYMSNLRDKWV---EGKLRKLRIIHAKTGKRDTSGGIDLGEQ 1082
DB 758 MTSYQDHMETROQMSAELMKERLAVLKEENDKKRAEKQKREKAKNKENGKVEG---814
QY 1083 HPLGTPTPGRRRRKGGSDYDDDDDDSD-----QGDEDDDEDEKEDQK 1129
DB 815 --LG-----KTDRKRLVKEPQVDTAEADMISAVKSRLLALQAKKERIQE---REM 864
QY 1130 GKTTDI CEDEDEGQAAVEELEKQIEKLSKQSQYRRKLPDASHLSRNVFGPDYRRR 1189
DB 865 VKLERQAEERIRKHAAAEK-----AFQGIATAKALVMRRTPDGTDRNHR 911
QY 1190 YWIL-PRCGGIFVGMESGEGLEIAKEREKKAESVQIKEMFETSG-----DSLNCST 1245
DB 912 YWLFSDVPGFLIE-----KGWVHDSIDYRPN 938
QY 1246 DHCEQKEDLKEKMTNLFLOKPGSFKLSKLEVAKMPPESEVMTPKPNACANGCTLSYQ 1305
DB 939 HHC-----KQHT-----VSGDEYCPRSK-----KANLGNA-----965
QY 1306 NSGRHSLGVSQSTATQSNVEKADSNLNTGSSGPKFYSPLPNDQLLTKLITEKNRQ--W 1363
DB 966 -----SMNTQHTATEVAE-----TTTPKQGNLW 991
QY 1364 PSLLPRTPCDDTSLTHADMSTASLVTQSOPPSKSPPTAPLGSQAQNPVGLNPFALSP 1423
DB 992 FL-----CD-----995
QY 1424 LQVKGVSMMGLQFCGWPTGVVTSNIPFTLSVPSLGLSGLSEGNNGSFLTSNVASKSE 1483
DB 996 -----995
QY 1484 SPVPQNEKATSAQAAVEAVAKPVPDPSPKPIPEMQFGWWRIIDPEDLKALLKVLHRLGI 1543
DB 996 -----SQKELDELNCLHPQGI 1012
QY 1544 REKALQKIQK-HLDYITQACLKNK-DVAIELNENEB--NQVTRDIVENWSVEQAMEM 1599
DB 1013 RESQLKERLEKRYODI IHSIHLAKPNLKGSCDGNLFLRSDLIE---VATRLQKG 1069
QY 1600 DLSVLQOVDELERVAS-----ASLQ-----VKGMWCPPEPASEREDLVYFEH 1641
DB 1070 GLGVETSEFARVILEKLDKDFECVIALQASVIKFFLQGGVAPKQ-----1117
QY 1642 KSFTKLCKEHDGFTGEDSSAHALERKSDNPLDIAVTRLADLERNIERRIEEDIAFGLR 1701
DB 1118 -----KRRKLQSEDSANTEVEDEKQWVE--EAKVASALE 1150
QY 1702 VWRRLSEARSAAQVALCIIQOLQKSIWEKSIKMYVCOI CRKGNBELLLLCDDCGDKGCH 1761
DB 1151 KWKTAREAQFTSRMVLGLMDACIKWMSAENARCKVCPKKGEDDKLILCDECNKAFF 1210
QY 1762 TYCHRPKITTIIPDGWFCPACIAKASQTLKIKLHVKKKTNSKGGKVTLTGDTDE 1821
```

```
Db 1211 LFCLRALYEVDPGEWQCACOPATA-----RRNSGRNNTESASB-----DSBDD 1257
Qy 1822 DS-----ASTSSSLKRGKNDLQKRKMBE 1844
Db 1258 ESEDESEEEEEEDYEVAGLRLPRKTIIRKHSVIPPAAARSRRPCKKPHSTRSQP 1317
Qy 1845 NTS--INLSKQSFTSVKPKRDKSDKDALCMLITEMETHEDAWPFLLPVNLKLVPGYK 1902
Db 1318 KAPPVDDAEVDLVLTQTKRSSRQSLQKCEIHLKIVKVFSPFPREPVTRDEADYY 1377
Qy 1903 KVIKPMDFSTIREKLSQYENLTFEALDVLVDFDNCSTEN 1944
Db 1378 DVITHPMDFTQVONKSCSGYSVQEFLETKMQVFTNAEYIN 1419

RESULT 35
US-10-702-148-27
; Sequence 27, Application US/10702148
; Publication NO. US20040063145A1
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042002
; CURRENT APPLICATION NUMBER: US/10/702,148
; PRIOR FILING DATE: 2003-11-05
; PRIOR APPLICATION NUMBER: 2001-04-20
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: US 09/418,710
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 27
; LENGTH: 1527
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-702-148-27

Query Match 5.2%; Score 532; DB 15; Length 1527;
Best Local Similarity 18.0%; Pred. No. 2.4e-17;
Matches 354; Conservative 255; Mismatches 597; Indels 756; Gaps 68;

Qy 189 KPLSLVNQAKTYMKLIVSPDVLKAGNKTSESSLLTSELRSKRGY-----KOA 241
Db 8 KPFLVNLPGBEFP-FTIPH-----TQFAFRTEYEARELERYSERIWTCKST 55
Qy 242 FPSQLKQESSKSLKVVAAALSNPKATSSSPA-HPKQTLNHNHNPFLTNALLGNHQPNG 300
Db 56 GSSQLTHKEAWEEGEVLAELK-----BEFFAWYKLVLEMVHN-----TA 97
Qy 301 VTQSVIQAPALATTKTKMQKINENIAAASSTPSPVNLSTSGRRTPGNQTPVMPAS 360
Db 98 SLEKLVDTAWLIMTKYAVECDEFV-----124
Qy 361 PTLHSQKKEKAVNNVNPVKTHQHSHPAKSLVEQFRGTDSIDIPSSKSDSESNDEEEDDE 420
Db 125 -----GKEKML--KVKIVKI--HPL-----EKVDEE 146
Qy 421 EDEDEDEDESDSQSDSNSESDTEGSEDEDDDDKQDSDSD-----T 467
Db 147 ATEKSDGACDPSDCKENSSQIAQDHQKKTETVWVEDEGRRESINDRARRSPKLPSTSLK 206
Qy 468 EGE-----KTSMKI-----NKTTSVKSPSKSLTGHSTPRNLHIATK-----503
Db 207 KGERKWAPPKFLPHKYIDVKLQNEKDIIINV---PAOSLIRTPPNKEIVYFIRNIALRA 264
Qy 504 -----APGSAPAALCESQSPA-----FLGTSSSTL---TSSPHSGTSKRRRTV 544
```

```
Db 265 GTGENAPVWVEDELVKYKSYLSPKSFSDFLDPYKYMTLNPSTKRKNTGSPDRKPSKKSK-T 323
Qy 545 DERELRIPLYG-WQRETRIRNFGG---RLOGEVAYYAPCGKKLRQYDEVIKYLSRNGIM 600
Db 324 DNSSLSPLNPKLWCHVHLKKSLSGSLPKVKNSKNSKSP-----EEHLEMMKMSPN--- 376
Qy 601 DISRDNFSFAKIRVGDIFYEARDGQEQMOWCLLKEEDVIPRAMEGRGRPPNDRQRA 660
Db 377 -----KLHT-NPHIPKKG-395
Qy 661 REESMRRRKGRPPNVGNAEFLDNADAKLLRLKLAQEIARQAQIKLLRLKQKQEOARVA 720
Db 396 KHSRPLKAKGRSKGI-----LNGQKSTGNS 421
Qy 721 KEAKKQQAIAAAEKKKQEQIKMKQOEKIKRIO-----IRMEKEL--RAQQIL 769
Db 422 KSPKK--GLKTPKTQKQWTLDDMAKGTQKMTAPRNSGGTPRTSSKPKHLPPAALHLI 479
Qy 770 EAKKKKKE-----AANAKLLEAEKRIK-EKEMR---ROQAVLLKHQER-----809
Db 480 AYYKENKDREDKRSALSVCVISTARTLLSSEDRARLPEELRSVLQRYELLEHKRWASMS 539
Qy 810 ERRQHMMLKAMEARKKAEKERLQEKREKRLNKERKLEQORLELEMAKELKFPNEED 869
Db 540 EQRKEYLKKREELKKKLEKAKERREKEMLERLEKQRYEDQEL---TGKYL-----590
Qy 870 MCLADQKPLPELPRIPGLVLSGTFSDCLMVVQVFNRFKVLGVDFVNDVNLVLOEGL 929
Db 591 -----PAFLVDTPE--GLPNTLFGDVAMVVEFLSCYSGLLLPDAQPITAVSLEALS 642
Qy 930 LNIQDSMGEVDLLVRLLSAAVCDPLITGYKAKTALGEHLNVLNVDNVDNSEIILQIFM- 988
Db 643 ADKGGFL-YLNRVLVILLQTLQDB-IAEDY---GELGKWLSEIPLTLHSVSELVRLCLR 697
Qy 989 -----EAHCGQTELTESLTKAFQAFQATPAQKASVLAFILNELA 1026
Db 698 RSDVOESESGSDTDDNDKDSAAFEDNEVDQEFLEKLETSEFFELTSEEKLIQILTALCHRL 757
Qy 1027 CSKSVWSIIDKNIDYMSNLRDQWV---EKLRLKLRIIHAKTGKRDTSGIDLGBEQ 1082
Db 758 MTYSVQDHMETRQMSAELWELRLAVLKEENDKKAERKQKREKEMAKKENGKENG-814
Qy 1083 HPLGTPTEPKRRRKGSDSDYDDDDDDDD-----QGDEDEDEDEKEDQK 1129
Db 815 --LG-----KTDRKKRIVKPEPQVDTAEADMISAVKSRLLAIQAKKEREIOE--REMK 864
Qy 1130 GKKTIDICEDEDEGDOAAASVELEKQIEKLSKQOQSYRRKLFDASHLSVSMFQPRYRR 1189
Db 865 VKLERQAEERIRKHAAAEK-----AFQEGIAKAKLVMRRTPTGTDNRNHR 911
Qy 1190 YWIL--PRCGGIFVEGMESGEGLEEIAKEREKLLKKAESVQIKEEMFETSG---DSLNCST 1245
Db 912 YWLFSDVEVPLGFIIE-----KGWVHDSIDYREN 938
Qy 1246 DHCEQEKDLKEDNTNLFQKPGSFKSLLEVAKMPPESEVMTPKPNAGANGCTLSYQ 1305
Db 939 HHC-----KDHT---VSGDEDYCPRSK-----KANLGKNA-----965
Qy 1306 NSGKHLSGVSQSTATQSNVEKADSNLNTGSSGPKGYSPLPNDQLLKTLEKTRQ--W 1363
Db 966 -----SMNTOGTATEVAVE-----TTTPKQGNLW 991
Qy 1364 FSLLPRTCDTSLTHADMSTASLVTPQSPKSPSPAPLGSQAQNPVGLNFPALSP 1423
Db 992 FL-----CD-----995
Qy 1424 LQVKGVSMMGLQFCWPTGVVTNIPFTLSVPSLGLGLSEGNGNSFLTNSNVASKSE 1483
Db 996 -----995
Qy 1484 SPVPQNEKATSAQPAAEVAKVPDPPSPKPIPEEMQFGWRIIDPEDLKALLKVLHLROI 1543
Db 1543 -----995
```

```
Db 996 -----SQKDELLELNLHPQGI 1012
Qy 1544 REKALQKQIQ-KHLDYITQACLKKNK-DVAIITELNENEB--NOVTRDIDVENWSVEEQAMEM 1599
Db 1013 RESQLKERLEKRYQDIHSHILARKPNLGLKSCDGNQELLFLRSDLLIE---VATRLQKG 1069
Qy 1600 DLSVLQOVEDLERRVAS-----ASLQ-----VKGWMCPEPASEREDLVYFEH 1641
Db 1070 GLGYVEETSEFEARVISLEKLKDFGECVIALQASVIKKFLOGFMAKQ----- 1117
Qy 1642 KSFTKLCKEHGDEFTGEDESSAHALERKSDNPLDIATVRLADLERNIERRIIEEDIAPGLR 1701
Db 1118 -----KRRKLOSEDSAKTEEVDDEKKMVE--EAKVASALE 1150
Qy 1702 VWRRLSEARSAAQVALCQIQLOKSIANEKSMKVVYCOICRKGDNEEILLLLCGDCDKCH 1761
Db 1151 KWKTAIREAQFTRSMHVLGLMDACIKWDMGAENARCKVCPKKGEDDKLILCDECNKAFH 1210
Qy 1762 TYCHRPKITITPDGDFPCPACIAKASGQTLKIKLHVKGKKTNSKKGKVVTLTGDTDE 1821
Db 1211 LFCLRPALYVEPDGEMQCPACOPATA-----RNSGRNYTESASE-----DSEDD 1257
Qy 1822 DS-----ASTSSSLKRGKNDLQKRKMBE 1844
Db 1258 ESDEEBEEBEEBEEDYEAGLRPRKTIIRKXHSVIPPAAARSGRRRPGKPHSTRSQP 1317
Qy 1845 NTS--INLSKQESFTSVKPKRDDSODLALCSMLITEMETHEDAWPFLLPVNLKLVPGYK 1902
Db 1318 KAPPVDDAEVDLVLOTKRSRROSLELOKCEEILHKIVKRYFRSFPFREPVRDEAEDYY 1377
Qy 1903 KVIKPMDFSTIREKLSGSGQPNLETFALDVLRLVFDNCETFN 1944
Db 1378 DVITHPMDQITVONKSCGSRVSRVQEFITDKMQVFTNAEVYN 1419

RESULT 36
US-09-864-761-37847
; Sequence 37847, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
```

```
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 37847
; LENGTH: 97
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC008277.1
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.81
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.83
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 0.95
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.74
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.89
; OTHER INFORMATION: SWISSPROT HIT: P16167, EVALU6 6.10e+00
; OTHER INFORMATION: EST_HUMAN HIT: AW754125.1, EVALU6 7.00e-25
US-09-864-761-37847

Query Match 4.8%; Score 491; DB 9; Length 97;
Best Local Similarity 100.0%; Pred No. 9.2e-17;
Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 439 SDSNSDTEGSEEDDDDDKQDESDSDTEGKTSMLKNTTSSVKGPSMLTGHSTPRN 498
Db 1 SDSNSDTEGSEEDDDDDKQDESDSDTEGKTSMLKNTTSSVKGPSMLTGHSTPRN 60

Qy 499 LHIKAPGSAAPALCSESQSPAPFLGTSSSLTSSPHS 535
Db 61 LHIKAPGSAAPALCSESQSPAPFLGTSSSLTSSPHS 97

RESULT 37
US-09-864-761-36754
; Sequence 36754, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
```



```
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 36754
; LENGTH: 91
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC008277.1
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.84
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.89
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 0.77
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 0.96
; OTHER INFORMATION: EST HUMAN HIT: AUI29635.1, EVALUATE 2.20e+00
; OTHER INFORMATION: SWISSPROT HIT: P13608, EVALUATE 9.60e-01
US-09-864-761-36754

Query Match 4.6%; Score 463; DB 9; Length 91;
Best Local Similarity 98.9%; Pred. No. 2e-15;
Matches 90; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1426 VKGGVSMGLQPCGPTGVVTSNIPFTLSVPSLGLSGLSEGNFSFLTNSVASSKSESP 1485
Db 1 VKGGVSMGLQPCGPTGVVTSNIPFTLSVPSLGLSGLSEGNFSFLTNSVASSKSESP 60

QY 1486 VPONEKATSAQPAAEVAKPVDPPSPKPIPE 1516
Db 61 VPONEKATSAQPAAEVAKPVDPPSPKPIPE 91

RESULT 38
US-09-839-479-64
; Sequence 64, Application US/09839479
; Publication No. US2002003779A1
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; CURRENT APPLICATION NUMBER: US/09/839,479
; CURRENT FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; NUMBER OF SEQ ID NOS: 91
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 65
; LENGTH: 175
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-376-537-65

Query Match 4.4%; Score 445.5; DB 15; Length 175;
Best Local Similarity 53.4%; Pred. No. 3e-14;
Matches 94; Conservative 31; Mismatches 50; Indels 1; Gaps 1;

QY 850 LEORLELEMAKELKKPNEDMCLADQKPLPRLPGLVLSGSTSDCLMVVQFURNFGK 909
Db 1 LEORLELEMAKELKKPNEDMCLADQKPLPRLPGLVLSGSTSDCLMVVQFURNFGK 60

QY 910 VLGFDPVNDVFNLSVLQEGLLNIGDSMGVQDQLLVRLLSAAVCDPGLITGYKAKTALGSH 969
Db 61 VLGFDPVNDVFNLSVLQEGLLNIGDSMGVQDQLLVRLLSAAVCDPGLITGYKAKTALGSH 120

QY 970 LLNVGNRDNVSEILQIFMEAHCGGTETSLTKFAQHAHTPAQKASVLAFINEL 1025
Db 121 VSEIPLTRDNVSEILRCPLMAYGVEPALCDRLRTQPPQAA-PPQKAAVLAFVPHL 175

RESULT 39
US-10-376-537-65
; Sequence 65, Application US/10376537
; Publication No. US20030224405A1
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; CURRENT APPLICATION NUMBER: US/10/376,537
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US/09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 65
; LENGTH: 175
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-376-537-65

Query Match 4.4%; Score 445.5; DB 15; Length 175;
Best Local Similarity 53.4%; Pred. No. 3e-14;
Matches 94; Conservative 31; Mismatches 50; Indels 1; Gaps 1;

QY 850 LEORLELEMAKELKKPNEDMCLADQKPLPRLPGLVLSGSTSDCLMVVQFURNFGK 909
Db 1 LEORLELEMAKELKKPNEDMCLADQKPLPRLPGLVLSGSTSDCLMVVQFURNFGK 60

QY 910 VLGFDPVNDVFNLSVLQEGLLNIGDSMGVQDQLLVRLLSAAVCDPGLITGYKAKTALGSH 969
Db 61 VLGFDPVNDVFNLSVLQEGLLNIGDSMGVQDQLLVRLLSAAVCDPGLITGYKAKTALGSH 120

QY 970 LLNVGNRDNVSEILQIFMEAHCGGTETSLTKFAQHAHTPAQKASVLAFINEL 1025
Db 121 VSEIPLTRDNVSEILRCPLMAYGVEPALCDRLRTQPPQAA-PPQKAAVLAFVPHL 175

RESULT 40
US-10-702-148-64
; Sequence 64, Application US/10702148
; Publication No. US20040063145A1
```

```
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 64
; LENGTH: 175
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-839-479-64

Query Match 4.4%; Score 445.5; DB 9; Length 175;
Best Local Similarity 53.4%; Pred. No. 3e-14;
Matches 94; Conservative 31; Mismatches 50; Indels 1; Gaps 1;

QY 850 LEORLELEMAKELKKPNEDMCLADQKPLPRLPGLVLSGSTSDCLMVVQFURNFGK 909
Db 1 LEORLELEMAKELKKPNEDMCLADQKPLPRLPGLVLSGSTSDCLMVVQFURNFGK 60

QY 910 VLGFDPVNDVFNLSVLQEGLLNIGDSMGVQDQLLVRLLSAAVCDPGLITGYKAKTALGSH 969
Db 61 VLGFDPVNDVFNLSVLQEGLLNIGDSMGVQDQLLVRLLSAAVCDPGLITGYKAKTALGSH 120

QY 970 LLNVGNRDNVSEILQIFMEAHCGGTETSLTKFAQHAHTPAQKASVLAFINEL 1025
Db 121 VSEIPLTRDNVSEILRCPLMAYGVEPALCDRLRTQPPQAA-PPQKAAVLAFVPHL 175

RESULT 39
US-10-376-537-65
; Sequence 65, Application US/10376537
; Publication No. US20030224405A1
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; CURRENT APPLICATION NUMBER: US/10/376,537
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US/09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 65
; LENGTH: 175
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-376-537-65

Query Match 4.4%; Score 445.5; DB 15; Length 175;
Best Local Similarity 53.4%; Pred. No. 3e-14;
Matches 94; Conservative 31; Mismatches 50; Indels 1; Gaps 1;

QY 850 LEORLELEMAKELKKPNEDMCLADQKPLPRLPGLVLSGSTSDCLMVVQFURNFGK 909
Db 1 LEORLELEMAKELKKPNEDMCLADQKPLPRLPGLVLSGSTSDCLMVVQFURNFGK 60

QY 910 VLGFDPVNDVFNLSVLQEGLLNIGDSMGVQDQLLVRLLSAAVCDPGLITGYKAKTALGSH 969
Db 61 VLGFDPVNDVFNLSVLQEGLLNIGDSMGVQDQLLVRLLSAAVCDPGLITGYKAKTALGSH 120

QY 970 LLNVGNRDNVSEILQIFMEAHCGGTETSLTKFAQHAHTPAQKASVLAFINEL 1025
Db 121 VSEIPLTRDNVSEILRCPLMAYGVEPALCDRLRTQPPQAA-PPQKAAVLAFVPHL 175

RESULT 40
US-10-702-148-64
; Sequence 64, Application US/10702148
; Publication No. US20040063145A1
```

```
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042002
; CURRENT APPLICATION NUMBER: US/10/702,148
; PRIORITY FILING DATE: 2003-11-05
; PRIOR APPLICATION NUMBER: PRIOR APPLICATION NUMBER: US/09/839,479
; PRIOR FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 64
; LENGTH: 175
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-702-148-64

Query Match      4.4%; Score 445.5; DB 15; Length 175;
Best Local Similarity 53.4%; Pred. No. 3e-14; Indels 1; Gaps 1;
Matches 94; Conservative 31; Mismatches 50;

QY 850 LEQRLELEMAKELKPNEDCLADQKPLPRLPGILVLSGSTFSDCLMVVQFLRNFGK 909
DB 1 LEQRKQQMIIEEMKKPTEDMCLTDHQLPDRFSRVPGLTLPGAFSDCLTIVEFLHSGK 60
QY 910 VLGFVNIDVNLVQLGSLNIGSMGEVDLLVRLLSAAVCDPGLITGYKAKTALGEH 969
DB 61 VLGFPAKDVPSLVLQBSGLGSLGCVQDILLVRLKAAHDPGFPSCYQSLKILGEK 120
QY 970 LLNVGVNDNVSEILQIPWEAHCGQTETLSLTKFAFOAHTPAOKASVLAFILNEL 1025
DB 121 VSEIPLTRDNNVSEILRCPLMAYGVEFALCDRLRTOPPOAQ-PPQKAAVLAFPPVHEL 175

RESULT 41
US-10-062-831-94
; Sequence 94, Application US/10062831
; Publication No. US20030105297A1
; GENERAL INFORMATION:
; APPLICANT: Steven M. Ruben, et al.
; TITLE OF INVENTION: 32 Human Secreted Proteins
; FILE REFERENCE: PZ006P1
; CURRENT APPLICATION NUMBER: US/10/062,831
; CURRENT FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: 09/690,454
; PRIOR FILING DATE: 1998-11-10
; PRIOR APPLICATION NUMBER: PCT/US98/10868
; PRIOR FILING DATE: May 28, 1998
; PRIOR APPLICATION NUMBER: 60/044,039
; PRIOR FILING DATE: May 30, 1997
; PRIOR APPLICATION NUMBER: 60/048,093
; PRIOR FILING DATE: May 30, 1997
; PRIOR APPLICATION NUMBER: 60/048,190
; PRIOR FILING DATE: May 30, 1997
; PRIOR APPLICATION NUMBER: 60/050,935
; PRIOR FILING DATE: May 30, 1997
; PRIOR APPLICATION NUMBER: 60/048,101
; PRIOR FILING DATE: May 30, 1997
; PRIOR APPLICATION NUMBER: 60/048,356
; PRIOR FILING DATE: May 30, 1997
; PRIOR APPLICATION NUMBER: 60/056,250
; PRIOR FILING DATE: August 29, 1997
; PRIOR APPLICATION NUMBER: 60/056,296
; PRIOR FILING DATE: August 29, 1997
; PRIOR APPLICATION NUMBER: 60/056,293
; PRIOR FILING DATE: August 29, 1997
; NUMBER OF SEQ ID NOS: 229
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 94
; LENGTH: 196
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-062-599-94

Query Match      4.0%; Score 406; DB 14; Length 196;
Best Local Similarity 39.5%; Pred. No. 2.8e-12;
Matches 83; Conservative 38; Mismatches 55; Indels 34; Gaps 5;

; NUMBER OF SEQ ID NOS: 229
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 94
; LENGTH: 196
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-062-831-94

Query Match      4.0%; Score 406; DB 14; Length 196;
Best Local Similarity 39.5%; Pred. No. 2.8e-12;
Matches 83; Conservative 38; Mismatches 55; Indels 34; Gaps 5;

; NUMBER OF SEQ ID NOS: 229
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 94
; LENGTH: 196
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-062-831-94

Query Match      4.0%; Score 406; DB 14; Length 196;
Best Local Similarity 39.5%; Pred. No. 2.8e-12;
Matches 83; Conservative 38; Mismatches 55; Indels 34; Gaps 5;
```



```
Db 1797 TSSTTSTISPAQKVMVAPISGVSVTGTTKVLTKVGSPTATVFOQKNFHTQFATVWKQG 1856
Qy 1274 -----SKLELV-----AKMPPSEVMTPKPNAGAGCT----- 1301
Db 1857 QNSGVVQVQKVLGIIIPSSGTSTQQTTSFQPRATVIRPNTSGSGGTTSNSQVITGP 1916
Qy 1302 -----LSYQNSGKHSL----- 1340
Db 1917 QIRPGMTVIRPFGQSTLGLKAIIRPVVMPVQGAPOQVMTQIIRGQPVSTAVSAPNTVST 1976
Qy 1319 ATQSNVEKADSNLNTGSSGP----- 1340
Db 1977 PQKSLTSTNTSSNTQSSASOPRQOQGVKLTMAQLTQGTGSGNQGLTVVIQOQGT 2036
Qy 1341 -----GRFYSPLPDNLLK-TLTKNQWFSLLP-RTPCDDTSLTHADMSTA----- 1385
Db 2037 TGQLQLIPQGVTVLPFGQOQQAAMPNGTVQRFLLTFLATATTATTTTSTVSTAAGT 2096
Qy 1386 -----SLVTPQSQ-----PPSKS-----PSTPAP----- 1405
Db 2097 GEQROSKLSPQVQVHODKTLPPAQSSVGPAPAKAQPTAQPAPQOPOTQOPSPAPQEVOT 2156
Qy 1406 -----LGSSAQ-----NPVGLNPPAL-SPLQVKGVSVM 1432
Db 2157 QPEVQTQTTVSSHVPSEAQPTHAQSSKQVAAQSQPQSNVQSQSPRVQSPQTRIREST 2216
Qy 1433 MGLQFCGHTGVTNSN---IPP---TLSVPSLG-----SGLGSGNGNSFTLS 1475
Db 2217 PSQISPGQOSQVQTTSPQIPQIPIQHTSLQIPSGQOPQSPQVQSSQSTQTLSSGQ-----TL 2271
Qy 1476 NVASSKSES-----PVPQNEKAT-----PQPAAEVA 1503
Db 2272 NOVSVSSPRPOLQIQOPQVQVIAVQLOQVQVLSQIQSVQVQIAQOQSGVQVQIKLQ 2331
Qy 1504 KPVDPPSPKPIPEMFGWRIIDPDLKALLVHLRGIREKALQKQIKHLDIYITQAC 1563
Db 2332 LPIQIQSSAVQTHQIQNVVTVQAAASVQEQVQQRVQQLRDQOQKKQKQIQIKREHTLQAS 2391
Qy 1564 LKN-----KDAVITE-----LNENEQVTRDIVENWSVEEQAMMDLSVLQ 1605
Db 2392 NOSEIIQOVVMKNAVIEHLKQKSMTPAREENQ--RMIVCN-----QVMKY-----ILD 2441
Qy 1606 QVEDLERRVASASLQVKGWMCPEPASEREDLVYFEHKSFTKLCKEHGDEFTGEDESSAHA 1665
Db 2442 KIDKEEQAQK-----KXKEESVE-QKRSKQVATKLSALLFKHKEQLRAEI 2487
Qy 1666 LERKSDNPLDIAVTRLADLERNIERIBEDI----- 1696
Db 2488 LKGRALLDKDLQIEVQEBELKRDLIKKEKDLMLQAATAVAACPVPVTPVLPAPPAPPPS 2547
Qy 1697 ---ARGLR-----VWRALSEARSAAQVALCQOQKSIWAKSIAWKYVQI 1740
Db 2548 PPPPGVQHTGLSTPTLPVASQKREKREEDSSSKSKKXMI STTSKETKDTKLYC-I 2606
Qy 1741 CRKGDNEELLLLDCDQKQCHYCHRPKITTPD-----GWFCPACIAKASGQTLKIKK 1795
Db 2607 KCTPYDESKFYI--GCDDR-QNWHGRCVGIQSEAEIIDEVCPQC----- 2650
Qy 1796 LHVKGKKTNESKKGKVTLTGDTDEDEDSASTSSSLKRGNKDLQKRMKEENTSINLSKQES 1855
Db 2651 -----QSTEDAMTVLTPTE----- 2665
Qy 1856 FTSVKKPKRDSKDLALCSMLITEMETHEDAWPFLLPVNLKLVQYKVKVKKPKMDFSTIR 1915
Db 2666 -----KQYEGKLRVLSIQAHKMAWPLEPVPDPNADPDYGVVKEPMDLATME 2713
Qy 1916 EKLSSGQPNLETALDVRLVFDNCETENEDSDSIIGRAGHNMRYKFEKKWTDTFKVS 1972
Db 2714 ERVQRYYEKLTETFEVADMTKIFDNCRYNPSDSPFYQCAEVLESFFVQK-LKGFAS 2769
```

RESULT 44

US-10-408-765A-254

```
; Sequence 254, Application US/10408765A
; Publication No. US20040101874A1
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fahy, Eoin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; APPLICANT: Warnock, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; FILE OF INVENTION: IDENTIFIED IN THE MITOCHONDRIAL PROTEOME
; FILE REFERENCE: 660088.465
; CURRENT APPLICATION NUMBER: US/10/408,765A
; CURRENT FILING DATE: 2003-04-04
; NUMBER OF SEQ ID NOS: 3077
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 254
; LENGTH: 3225
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-408-765A-254
```

```
Query Match 3.7%; Score 372.5; DB 16; Length 3225;
Best Local Similarity 18.5%; Pred. No. 3.2e-09;
Matches 402; Conservative 327; Mismatches 779; Indels 665; Gaps 84;
```

```
Qy 19 EQSKNQPLDARDVKIKKKPR--KKAMESSNSDSGTSTSDTSSGSSDSDDDLEDE 76
Db 762 EQIHSLSTEA---KSKDKIEVLQNELDDVQLQSEQSTLIRSLQSQIQNKSEVLEGA 818
Qy 77 EEDQSTEEESDDSDSESAQHSNNQVLLHGISDPKADGQKATEKAQEKRIHQPLA 136
Db 819 RVRISSKVELSQALSKQELIETKMDLLEKRDVETLQQTTEEDKQ-----QVTEIS 873
Qy 137 FESQTHFSQSQK-----QPVLSQLPFTFQSSQAKESVKNKHTSVIQSTGLVSNVKP 190
Db 874 FSWTEKVVQVNEEFKSLGVEIKTLKEQLNLLSRAEAKKEQVEEDNEV--SSGLKQNYDE 931
Qy 191 LSLVNAQKETYMKLIWPSPDVLKAGN-----KNTSESSLLTSELSKRE 236
Db 932 MSPAQIQSKELQHEF---DLLKKEQRRKQLQAALINRKELLQVRSLREELANLKD 987
Qy 237 QYKQAFP---SQLKQKSSLSKVIKVAALSNPKATSS---PAHPKQTL-----ENNHP 284
Db 988 ESKKEIPLSETERGEVEEDKENKEY-----SEKCVTSKQBIYILKQTIIEKEVELOHI 1042
Qy 285 NPFLTNALGNHQPNQVGIQSVIQEAPLALTTKTK---MOSKINEN----- 326
Db 1043 RKDLEEKLAABEQ---FQALVKQMNQTLQDKTNQIDLLQAEISENQAIQKLTISNTDA 1098
Qy 327 -----IAAASSTPPSSPVNLSTSGRRTPGNTQVWPSASPIHSGQKEKAVSNVNVPKT 381
Db 1099 SDGDSVALVKTETVISPCTGSSSEHWKPELE-----EKILALEKEKEQKQKLEALT 1151
Qy 382 QHHSHPAKSLVEQRCGTDSQIPSSKQSDSDNEDEDEDEDEDEDEDD-----E 431
Db 1152 SRKA-ILKKAQEKERHLEELKQKQD--DYNRLQEQFDEOSKENENIGDQLRQIQVRE 1208
Qy 432 SDDSQSSESDNSE--SDTEGSEE-----EDDDDKQKQDES----- 464
Db 1209 SIDGKLSTDOQESCSSTPGLEEPLFKATEQHHTQPVLESNLCPDWPSHSEDASALOGGT 1268
Qy 465 -----SDTEGSEKT--SMKLNKTTSSVKSPS---MSLTGHSTPNLNLHTAKAPGSAPA 510
Db 1269 SVAQIKAQKLEIAEKVELELKVSSSTTSELTKKSEEVFQEQIQKQKQKLEISLKTVSHE 1328
Qy 511 ALCSSESQAPFLGTSSSTLSTSSPHSGTSSKRRVTDDEBELRIPLEYGVQWRETRIRNFGRL 570
Db 1329 AEVHAESLQKLESSQIQIAGLEHL-----RELQPKLD-----EL 1363
Qy 571 QGEVAYYAPCKKLRQYPEVVIKYLNRNGIMDISRDNFSFSAKIRVGDGFYEARDGPFQEMQW 630
```

Db 1364 Q-----KLISKKEEDVSYS-----GQISEKEAALTKIQT 1393
Qy 631 CULLKEEDVPIPRAMEGRGRPPNDQRARBEESRRRKGRPPVGNAAFLDNADAKIL 690
Db 1394 ETIEQEDLIKALHTQEMQAK--EHDRIKQLOVELCEMKQKPEBIEE--SRAKQIQI 1448
Qy 691 RKLQAEIARQAQIKLRLKQKQSQ-----ARVAKEAKQQAAMAAEKEKQKEQIKI 744
Db 1449 RKLQALISRKBA--LKENKSLQEEISLARGTITRIKTSIADVESQVSAQNKEKOTVGLRL 1507
Qy 745 -MKQKEKRIQIIRMEKELRAQQL-----BAKKKKKEEAANAKIL 785
Db 1508 ALQOERDKLITE--MDRSLLENQSLSSCESLKLALGLTEDKEKLVKEIETLSKSKIA 1565
Qy 786 EA-EKRIKEKEMRROQAVLLKHQ-----RERRRQHM-----LMKAMBARAKA 828
Db 1566 ESTEQEKHKELOKEVEILLQSYENVNSNEAERIQHVVEARQEKQELVKLRSTEAANK 1625
Qy 829 EKERLQEKREKELINKERKL-----EQRRLEMAKELKKPNEDMCLADQKPLPELPI 884
Db 1626 TEKLOEAEQEMEEKMKRPAKSKQKQKILE-----EENDRLRAEVHP----- 1671
Qy 885 PGLVLSGTSFSDCLMVVQFLRNFVGKVLGFDVNDVNLVLOEGLINTGDSMG-EVQDLL 943
Db 1672 -----AGDTAKECMETILSSNASKBEELRVQMEYETLSKFKQSLMSKDSISEVQDU- 1725
Qy 944 VLLLSAAVCDPGLITGYKAKTALGHELLNVGNRDNVSEILQIFMEAHCGQTTELTSKT 1003
Db 1726 -----KHQIEGNVSKQANLEA-----TEKHDNQINVTIEE-GT 1756
Qy 1004 KAFQHTPAQKASVLAFLINELACSKSVVSEIDKI-----IDVMSNLRDRKVVVEGKRL 1059
Db 1757 OSIPEGETEEDSLMS--TRPTCESVPSAKSANPAVSKDFSSHDEINNYLQIQDQKE 1813
Qy 1060 RIHAHKTGKRDTSGIDIGEOHPLGTPTPKERRRKGDSYDDDDDDSDQDQDEDD 1119
Db 1814 RIAGLEBKQKMKFSQTLNENKNTLS-----QISTKGELKMLQEEVTKNLLNQIQ 1869
Qy 1120 ED-----BEDKEDQKGTDTICEDEDEGQAAVELEKQIEKLSKQSQYRRK 1168
Db 1869 EELSRVTKLKEPAEKEKDDLE-----ERLMNQLAELNGSIGNYQD 1909
Qy 1169 LFDA---SHLSRVNFGPRYRRYWIILPRCGIFVEGWESGEGLEEIAKREKUKKAB 1225
Db 1910 VTDAQIKNELLESEMN-----LKKC-----VSELE--EERQQLVKEKTKV---ES 1950
Qy 1226 VOIKEEMFETSDSLNCNTDHC--BQKEDLKEKONTNLFLOKP-----GSFSKSLKL 1276
Db 1951 EIRKEYLEKIQQAQEPGNKSHAKELQELLKEKQEVQLOKDCIRYQEKISALERTYKA 2010
Qy 1277 LEVAKMPPSEVMTPKPNAGANGCTLSYQNSGKSHLSGS--VOSTATQSNVKEKADSNLNF 1334
Db 2011 LEFVQTESQKDLITKENLAQ---AVEHKKQAQELASFVKVLLDDTQSEAAVLADNL-- 2065
Qy 1335 TGSSGFGFYPLPNDQLLTKTEKNRWFSLLPRTPCDDTSLTHADNSTASLVTPQSQP 1394
Db 2066 -----KLKKELOQNSKESVK-----SQMKQKQDEDLERLRFQAEK 2099
Qy 1395 PSKSPSPAPIGSSAQNPVGLNPALSPLOV---KGGVSMGLQFCGWPTGVVTSNTPF 1451
Db 2100 HLKEKNKQEKLDALUREKVLHEE--TIGEIQTLLNKDKQEVQLOQ----- 2143
Qy 1452 TILVPSLGLSLSGNSFLTSNVASKSBSPPQNEKATSAQPAAVEAKPVDPPSP 1511
Db 2144 -----ENDLSTVTQLAAFTKSNSSLSQDDR----- 2168
Qy 1512 KPIPEMFGWRIIDPDLKALLKVLHLRGIREKALQIQKHLDYITQACLKNKQVAI 1571
Db 2169 -----RVID-----EAKWKERKFSQ-----AIQSKEE-- 2190
Qy 1572 IELNENEEN-QVTRDIVENWSVEQAMEMDLSVLQOVEDLERRVASASLOVKGWMC----- 1626
Db 2191 -EIRLKDNCVSLQDLQRMHSIHMEELKINISRLHDKQIWESKAQTEVQLQKQVCDTLQ 2249

Qy 1627 ---PEPASEREDLVYFEHKHSFTKL-----CKEHGDEFTG--- 1657
Db 2250 GENKELLQLEBTRHLHYHSSQNELAKLESELKSLKADQLTDLNSLSLEKCKEQKGNLEGIIR 2309
Qy 1658 ---EDESASAHALERKSDNPLDIATVTRLADLE--RNIERRTEEDIAPLRVVRRALSEAR 1711
Db 2310 QOERADIQNSKFSYBOLE-----TDLOASRELTSLRHLHEEINMKEQKIISLLSGKB 2358
Qy 1712 SAAQVALC--TQQLQKSIWEKSIKMKVYQCIQCRKGDNEELLLLLCDGCDKGCHTYCHRPKI 1769
Db 2359 EAIQVIAELAQOHDKEI---KELENLSQ-----EENENIVLEENKK--- 2399
Qy 1770 TTIPDGFWCAPACIAKASG--QTLK-IKKLHVKGK-----KTNESKKGKVVLTUTGDT 1819
Db 2400 -----AVDKTNQLMETLTKIKKENIQQAQLDSFVKSMSSQLNDRDRIVGDYQ 2447
Qy 1820 D-----EDSASTSSSLK--RG----- 1833
Db 2448 QLEERHLSIILEKQOLIQEAAANNKLBKBEIRGLRSHMDDLNSENAKLDAELIQYREDLN 2507
Qy 1834 -----NKDLQKR-----KMEENTSINLSKQESFTSVKPKCRDSS 1867
Db 2508 QVITIKDSQKQLLEVQLOQNKELKENYAKLEELKKESEANEDLRRSFNALQEEKQDLS 2567
Qy 1868 KDLALCSMLITEM 1880
Db 2568 KEIESLKVSISQL 2580

RESULT 45

US-09-864-761-33806
; Sequence 33806, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aemica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670

```
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 33806
; LENGTH: 70
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC008277.1
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.95
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.9
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.99
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.7
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.6
; OTHER INFORMATION: EST_HUMAN HIT: BF057266.1, EVALUATE 1.00e-24
; OTHER INFORMATION: EST_HUMAN HIT: A1339206.1, EVALUATE 8.00e-25
; OTHER INFORMATION: SWISSPROT HIT: P40640, EVALUATE 4.50e+00
US-09-864-761-33806

Query Match          3.6%; Score 363; DB 9; Length 70;
Best Local Similarity 100.0%; Pred. No. 1e-10;
Matches 69; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 628 MOWCLLKEDVIPRIAMEGRGRPPNPDQRAREESMRKRKGRPPNVGNAEFLDNADA 687
Db 2 MOWCLLKEDVIPRIAMEGRGRPPNPDQRAREESMRKRKGRPPNVGNAEFLDNADA 61

QY 688 KLLRKLQQAQ 696
Db 62 KLLRKLQQAQ 70

RESULT 46
US-10-144-198-2
; Sequence 2, Application US/10144198
; Publication No. US20030219748A1
; GENERAL INFORMATION:
; APPLICANT: Origene Technologies Inc
; TITLE OF INVENTION: Regulated Prostate Cance Genes
; FILE REFERENCE: 9U 105 R1
; CURRENT APPLICATION NUMBER: US/10/144,198
; CURRENT FILING DATE: 2002-05-14
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 1564
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-144-198-2

Query Match          3.5%; Score 357; DB 15; Length 1564;
Best Local Similarity 18.5%; Pred. No. 7.8e-09;
Matches 342; Conservative 255; Mismatches 68; Indels 568; Gaps 70;

QY 8 SSGGNRKCNOEQSKNQPLDARVDKIKDKPKRKAMSSSSDSGTSSTSEGISSS 67
Db 64 SPRKGYSNRYRSPERTGDLRRMKNKQDVDTPEQKRNTEBSSSPVRKSSRG-RHR 122

QY 68 DSDDL-----EDEDDEDSIE-ESEDDDSSESAQHKSNNQVLLHGISDPKADQKATE 122
Db 123 EKEDIKITKPTPESEENVETNRDSDN-----GDINYD----- 159

QY 123 KAOEKRIHPLAPESQTHSFOSQKQKQPVLSLQQLPFIQSSQAKESVNKHTSVIGST 182
```

```
Db 160 -----YVHLSLEMKRQKIQRELM-----KLEQENMEKREII--- 192
QY 183 GLVSNVPLSLVNOAKKETYNKLIVPSPDVLKAGNKVNTSESSLLTSELRSKRQYKQAF 242
Db 193 -----IKKEVSPVVRSKLSPSPS-LRKSCKSPKRKSPKSSS-ASKDKRTSAV 240
QY 243 PSQLKQOE-----SSKSLKKVIAALSNPKNATSSSPAHPKQ-TLENNHPNPFLTNALLGNHQ 297
Db 241 SSPLLDQQRNSKTNQSKKK-----GRTSPSPPIPEDIALGKKYKVKVDRIEKT 294
QY 298 PNG-----VIQSVIOEAPIALATTKTMQSKINENIAAASSTPFPSPVNLSTGRR--- 347
Db 295 RDGKDRGRDFERQREKDKPRSTSPAGQHSPISRRHHSSSSQSGSSIQRHSPPRKRT 354
QY 348 -TPGNQTPVMP-----SASPI-----LHSGQKEKAVNNVNPVKTO-HHSHPAKSLVEQPRG 397
Db 355 PPSYQRTLTPLPLRRSASYPVSHLSLPQRKSPRHRSPMRKGRHDHRTS----- 407
QY 398 TDSIPSSKOSEDNSNEDEED-----DEEDEDEDEDESDSSESNSSESDEGSEE 452
Db 408 -----QSHDRRHERREDTRGKDRKDSREEREYEQQSSSRDHRDREPRDGRDR 458
QY 453 EDDDDKQDESDDTEGEKTSMLKNTTSSVKSPPSMLTGHSTPRNLHIAKAPGSAPAL 512
Db 459 RDARDTRDRRLRDSRDMRDSREMDYSRDTK----- 490
QY 513 CSSESQPAFLGTSSSTLTSSPHSGTSKRRRVTDRELRIPLE-----YGNQR-----ETRI 563
Db 491 --ESRDP-----RDSRSTRDAHYDRDREGDRTHRKEDTYPESRSYGRNHLRSESSRTEI 543
QY 564 RNFG-----GRLOGEVAYYAPCGKKLRQYPEVIKYLRSNGIMDISRDNSFS 610
Db 544 RNESRSRSEIRNDRMGRSGRVP-----ELPEKSGRSRGSQID-----SHS 587
QY 611 AKIRVGDFYARDCGPQBMQWCLLKEDVIPRIAMEGRGRPPNPDQRAREESMRKRK 670
Db 588 SNSNYHDSWETRSTYPE-----RDYPERDNRDQARDSFERRH 626
QY 671 GRPPNVGNABFLDNADAKLLRKLQQAQETARQAQAIKLLRKLQKQEQARVAKEAKKQQAIM 730
Db 627 GERDRDRNRERDQRPSPPIRHOGRNDELERD-----ERREERRVDVDRDRDE-- 674
QY 731 AAEEKRKQKQIKIMKQOEKIKRIQOIRMEKELPAQOILEAKKKKKEAANAUKLEAE-- 788
Db 675 RARERDRERDRERERERERERERERERERERERERERERERERERERERERERERER 734
QY 789 -----KRIKEKEMRRQQAALLKHQERERRRRQHMMLKAMEARKKAEKERL-----KQ 836
Db 735 RERERERERERERERERERERERERERERERERERERERERERERERERERERERER 794
QY 837 EKRDEKRLNKERKLEQRRL---ELEMAKELKKPNEDMCLADQKPLPRLPTIPGLVLSGST 893
Db 795 KGRDRREKREIREDRNPDRGHDHDKSKKRYRNEGSPSPRQSPKRRREHSPD----- 847
QY 894 PSDCLMVVQFLRNFQKVLGPDVNDVNLVLQGLLNIQDSMGVEVDLLVRLLSAAVCD 953
Db 848 -SDA-----YNSGDDKNEKH-----RLLSQVY-- 868
QY 954 PGLITGYKAKTALGEHLNLVGNVRDNLVSEIIQIPMEAHCGQTELTSJTKT-AFQAHTPA 1012
Db 869 -----RPOESRSL-----SPSHLTEDRQGRWKEEDRKPE 897
QY 1013 QKASVLAFLINELACSKSVVSEIDKNIDYMSNLRRDKWVE-GKLRKLRIIHAKTGKRD 1071
Db 898 RKESRRAYEEQEL---KEKVSSVDKQ-----REQTEILLESSRMRAQDIIGHHQSDE 947
QY 1072 TSGGIDIGEGHQPLGTTPGKRKRKGGDSYDDDDDDDDDDQDQD-----EDDEDEDEKE 1126
Db 948 TS---DRAHDENKKKAKIQKKPIKK-----KEDDVGIERNIETTSDEGQVFSPPK 996
QY 1127 DQKGKTKTDICEDDEGQQAASVEELEKQIEKLKSKQSQYRKLFDASHLSRNVFPGDPRY 1186
```



```
Db 997 GQKKSIEKRRKSKGSDISDEE-----AAQSKKKG- 1030
Qy 1187 RRRYWLPCGGI-----FVEGMESGEGLEIEAKEREKLLKKAESVQIKKEEMFETSGDSL 1241
Db 1031 -----PRTPPTTKEELVENCNGNGILEDQKKE-----DTAFSDMSD--- 1069
Qy 1242 CSNTHCEQKEDLKEKDNVNLFLQPGSP-SKSLKLELVAKMPPSEVMTTPNAGAGCC 1300
Db 1070 -----EDVDPRTVEAHT-ATATTGSPSPSLSL-----PP-PPVATATAT 1113
Qy 1301 TL-----SVQSGKSHLSGVOSTATQSNVERKADSNLENLSSGSGKGYSPLPNDQ 1351
Db 1114 TYPATLAATAAATSFSAITSTATPTW-----TNTPT----- 1151
Qy 1352 LLKTLTEKNRWFSLPRTPCDDTSLTHADMSTASLVTPSQSPKSPSPPTAPLGSSAQ 1411
Db 1152 -----ANEDSHRKCHRTVEKVPHTVETIEDAQHRKPMQKRS- -LGSNRS 1197
Qy 1412 NPVGLNPFALPLQ---VKGVSMWG---LQFCGPTGVVTNIPFTLSVPSLGSGLGLS 1465
Db 1198 NRSHTSGRLRSPNSAHRSGDQSGRKVLHSGSRDRREKTSLEITGERKSRIDQLKRG 1257
Qy 1466 EGNNSFLTNSVASSKSESPVQNEKATSAQAPAAVEAKPV-----DPSPKPIPEE--- 1517
Db 1258 E-----PKRSTSSDQSRKSHSRKSPESDRQVHRSRSGSPDRRLQERDRY 1305
Qy 1518 -----MFGW-----W-----RIIDPDLLKALLKVLHRLGIREKALQK 1550
Db 1306 EHRERERERRTRQEWDRDADKWPNNRDRDLRERERERERERERERERERERERERER 1365
Qy 1551 QIQKHLDYTOCLKNKQVAIIELNENE-ENQVTRDIV-----ENWSVEEQAMEMDLVLQ 1605
Db 1366 SVDRDRDRDRFTFESSQIESVKRCEAKLEGEHERDLESTSRDSLALDKERMDKLGVSQ 1425
Qy 1606 QVEDLERRVASLQVKGWMCPEPASEREDLVYFPHKSFCLKCKEHDGEFTGEDSES---S 1662
Db 1426 GFEDTNKSRTESSLE-----S-----AGDDESKLDD 1450
Qy 1663 AHAL-----ERKSDNPLDIATRLADLERNIERRIEEDIAPGLRV 1702
Db 1451 AHSLSGAGEGYEPISDDELDILA--GDAEREDQODEKMPDLDV 1496
```

RESULT 47

```
US-10-144-198-4
; Sequence 4, Application US/10144198
; Publication No. US20030219748A1
; GENERAL INFORMATION:
; APPLICANT: Origene Technologies Inc
; TITLE OF INVENTION: Regulated Prostate Cance Genes
; FILE REFERENCE: 9U 105 R1
; CURRENT APPLICATION NUMBER: US/10/144,198
; CURRENT FILING DATE: 2002-05-14
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 4
; LENGTH: 1564
; TYPE: PRN
; ORGANISM: Homo sapiens
US-10-144-198-4
```

```
Query Match 3.5%; Score 357; DB 15; Length 1564;
Best Local Similarity 18.5%; Pred. No. 7,8e-09;
Matches 342; Conservative 255; Mismatches 683; Indels 568; Gaps 70;
```

```
Qy 8 SSGGNRKNQKQKQNPDLARVDKIKDKKPKKKAEMSSNSDSGTSSTSSGSISS 67
Db 64 SPRKGYSNRYRSPERTGDLREMKNRQVDVTEPKRNTTESSSPVRKSSRG-RHR 122
Qy 68 DSDDL-----EEDDEEDQSI-ESEDDSDSESAQHKSNQVLLHGISDPKADQKATE 122
Db 123 EKEDIKTKERTPESEENVEWETNRDSDN-----GDINYD----- 159
```

```
Qy 123 KAQEKRIHQIPLAPESQTHSFQSQOKQFQVLSQQLPFIQSSQAKESVNRKHTSVIOST 182
Db 160 -----YVHELSELMKQKIQRELM-----KLEQENMEKREBII--- 192
Qy 183 GLVSNVPLSLVNOAKKTYMKLIIVPSPDVFLKAGNKNTSESSLLTSELRSREQYKQAF 242
Db 193 -----IKKEVSPPEVVRSKLSPSPS-LRKSSKSPKRKSSPSKSSS-ASKDKRTSAV 240
Qy 243 PSQLKKQES-----SSSLAKKVIAAALSNPKATSSSPAHPKQ-TLENNHPNPLFNALLNHQ 297
Db 241 SPSLLDQQRNSKTNQSKK-----GPRTPSPPPPIPEDIALGKVKYKVKYKDRIEKT 294
Qy 298 PNG-----VIQSVIQEAPALATYTKQKSKINENIAAASSTPSSPNVLTSGRR--- 347
Db 295 RDGKDRGRDFERQREKDRKPRSTPAQOHHSPISSRHHSSSSQSGSSIQRHSPSPRRKRT 354
Qy 348 -TPGNQTPMP-----SASPI-----LHSQKEKAVSNVNPVKTO-HHSHPAKSLVEQFRG 397
Db 355 PPSYQRTLTTPLRKSASPYSPSHLSLSPQRKQSPPRHSPMKRKGRRDHRTS----- 407
Qy 398 TDSPTSSKDESDSNEDEED-----DEEEDDEDDDDSDSQSDESNSSESTEGSBE 452
Db 408 -----QSHDRRHERREDTRGKRDREKDSREEREYEQDOSSRDRDRDRDRDRDR 458
Qy 453 EDDDDKQODESDSOTEGEKTSMKLNKTTSSVKSPMSLTGHSHTPRNLHIAKAPGSAAL 512
Db 459 RDARDTRRRLSRDRMRDSREMRDYSRDTK----- 490
Qy 513 CSSESQPAFLGTSSSTLTSSPHSGTSKRRRTVDERELRIPLE---YGHQR-----ETPI 563
Db 491 -----ESRDP-----RDSRSRTRDAHYDRDREGDRTHRKEDTYPEESRYGRNHLREESRTEI 543
Qy 564 RNFG-----GRLQGEVAYYAPCGKLRQYPEVIKVLSRNGIMDISRDNFSPS 610
Db 544 RNESNESRSRIRNDRMGRSRGPV-----ELPEKSGRSGRGQID-----SHS 587
Qy 611 AKIRVGDFYEARDGPQEMQWCLLKEEDVIPRAMEGRGRPPNPDRQARAESRMRRK 670
Db 588 SNSVHDSWETRSTTYPE-----RDRYPERDNDRDQARDSFFERRH 626
Qy 671 GRPPNVGNAEFLDNADAKLLRLKLAQEIARQAQIKLLRKLOKQOQAVAKAKQQAIM 730
Db 627 GERDRDRNRDRQRPSPPIRHQGRNDELERD-----ERREERVRDVRDDRDE-- 674
Qy 731 AAEEKRKQEKQIKMKQOEKIKRIQOIRMEKELRAQOILEAKKKKEEAANAALLKLEAB-- 788
Db 675 RAREDRERDRERERERERERERERERERERERERERERERERERERERERERERERER 734
Qy 789 -----KRIKEKEMRQOAVLLKHQERERRRQHMMLKAMEARKKABEKERL-----KQ 836
Db 735 RERERERERERERERERERERERERERERERERERERERERERERERERERERERER 794
Qy 837 EKROEKRLNKERKLEORL-----ELEMAKELKPNEDMCLADQKPLPELPRIPGLVLSGT 893
Db 795 KGRDDREREREIREDRNPDRGDHDERKSKRYRNEGSPSPRQSPKRRREHSPD----- 847
Qy 894 FSDCLMVVQFLRNFQKVLGFDVNDIVPNLSVLQEGLLNIGDSMGVEQDILLVLLLSAAVCD 953
Db 848 -SDA-----YNSGDDKKNKX-----RLLSQV-- 868
Qy 954 PGLITGYKAKTALGEHLNVLNVGRNDRNVSEIIQIFMEAHCGQTELTSLTKT-AFAQHTPA 1012
Db 869 -----RPQESRSL-----SPSHLTEDRQGRWKBEDRKPE 897
Qy 1013 QKASVLAFLINELACSKSVSEIDKNIDYMNLRDKWVVE-GKURKURIIHAKTGTGRD 1071
Db 898 RKESRRRYEEQEL---KEKVSSVDKQ-----REQTEILESSRMRADQIIGHQSEDR 947
Qy 1072 TSGGIDLGEEOHPLGTPPGKRRRKGSDSYDDDDDDDDDDQGD-----EDDEDEEKE 1126
Db 948 TS---DRAHDKKKKAKIQKKPIKKK-----KEDDVGIERNIETTSQGVFSPKK 996
Qy 1127 DQKGGKTDICEDDEGQQAASVEELEKQIEKLSKQSQYRRRKLFDASHLSLRSVMFGPORY 1186
```



```
QY 1274 SKLLEVAKMPPESEVMTKPNAG 1296
Db 956 STHEQIASSPPGDNTPDDDPQG 978

RESULT 49
US-10-294-804-2
; Sequence 2, Application US/10294804
; Publication No. US20030133948A1
; GENERAL INFORMATION:
; APPLICANT: Robertson, Erle S.
; APPLICANT: Cotter, Murray A.
; TITLE OF INVENTION: Methods to Inhibit or Enhance the Binding of Viral DNA
; FILE OF INVENTION: to Genomic Host DNA
; CURRENT APPLICATION NUMBER: US/10/294,804
; CURRENT FILING DATE: 2002-11-14
; PRIOR APPLICATION NUMBER: US/09/410,399
; PRIOR FILING DATE: 1999-10-01
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 1162
; TYPE: PRT
; ORGANISM: Kaposi's sarcoma-associated herpesvirus
US-10-294-804-2

Query Match 3.2%; Score 326.5; DB 14; Length 1162;
Best Local Similarity 16.9%; Pred. No. 1.6e-07;
Matches 196; Conservative 191; Mismatches 405; Indels 371; Gaps 31;

QY 207 VPSDVLKAGNNTSESSLLTSELRSKREQYKQAPPQOLKQESSKSLKVVIALSNPK 266
Db 114 IPPSPVFGTDTISPAL-PTQSPSSQRPPLSPTGPDSTMPRP-----pp 165

QY 267 ATSSPAH-----PKOTLENNHPNPLNAL-----LGNHPNGVQSVIQEAPLALT 314
Db 166 SQQTPPHSPPTPPPEPPSKSPDSLAPSLARLKRLLSSPQSPSTLNPICQSPPV--- 222

QY 315 TTKMQSKINENIAAASSTP---FSPVNL-STSGRRPGNTPMP-----SASPLHSQ 366
Db 223 -----SPRCDFANRSVYPWATESPIYVGSSDGTTPRPPTSPISIGSSSPSEGSW 276

QY 367 G-----KEKAVSNVNPVKTQHSH 387
Db 277 GDDTAMVLVLAIEASKEKCESENQAGDNGDNISKEQVQDKDNDKDDDEEQE 336

QY 388 AKSLVEQFRGTDSDIPSSKSDSDNEDEDEDEDEDD-----EDDESDSQSESDN 442
Db 337 TD---EEDDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDE 393

QY 443 SESDFEGSEEDDDKQDESDDSTEGEKTSMKL-----NKTSSVKSPSLTGHSTPR 497
Db 394 EEEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDE 452

QY 498 NLHIAPGSAAPALCSQSAPFLGTSSSTLSSPHSSTSKRRVTRDERELRILEYGM 557
Db 453 -----QQQEPQ-----QQEPQEQEQEQEQEQEQEQEQEQEQEQEQEQEQEQ 486

QY 558 QRETRIRNFGGLQGEVAYAPCGKKLQYPIVKYLSRNGIMDTSRDNFSFSAKIRVGD 617
Db 487 QQEPQ-----QQE-----PQEPQEQEQEQEQEQEQEQEQEQEQEQEQEQEQEQ 505

QY 618 FYEARDGQEQWMLCKLEDVIPRTRAMEGRGRPNPDQR-AREESRMRRKGRPNV 676
Db 506 ----QQEPQEQEQEQEQEQEQEQEQEQEQEQEQEQEQEQEQEQEQEQEQEQEQEQ 561

QY 677 GNAEFLDNADAKLLKLAQETARQAQIKLLRLKQEQARVAKEAKKQQAIAAEKR 736
Db 562 REPQ-----QREPQEQEQEQEQEQEQEQEQEQEQEQEQEQEQEQEQEQEQEQEQ 610

QY 737 KQEQIKMKQEQEKIKRIQIRMEKELRAQILEAK--KKKEAANAALLAEKRIKEX 794
```

```
RESULT 50
US-10-052-648A-52
; Sequence 52, Application US/10052648A
; Publication No. US20040005558A1
; GENERAL INFORMATION:
; APPLICANT: Anderson, David
; APPLICANT: Burgess, Catherine
; APPLICANT: Casman, Stacie
; APPLICANT: Colman, Steven
; APPLICANT: Edinger, Shomit R.
; APPLICANT: Ellerman, Karen
; APPLICANT: Gerlach, Valerie
; APPLICANT: Gunther, Erik
; APPLICANT: Kekuda, Rameesh
; APPLICANT: MacDougall, John R.
; APPLICANT: Mehra, Meera
; APPLICANT: Paturajan, Meera
; APPLICANT: Rothenberg, Mark
; APPLICANT: Shimkets, Richard
; APPLICANT: Smithson, Glennda
; APPLICANT: Spytsek, Kimberly A.
; APPLICANT: Stone, David J.
; APPLICANT: Vernet, Corine A.M.
; APPLICANT: Zethusen, Bryan D.
; TITLE OF INVENTION: PROTEINS, POLYNUCLEOTIDES ENCODING THEM AND METHODS OF
; FILE OF INVENTION: USING THE SAME
; FILE REFERENCE: 21402-250 (CURA-550)
; CURRENT APPLICATION NUMBER: US/10/052,648A
; CURRENT FILING DATE: 2002-12-09
; PRIOR APPLICATION NUMBER: 60/262,454
; PRIOR FILING DATE: 2001-01-18
; PRIOR APPLICATION NUMBER: 60/272,920
; PRIOR FILING DATE: 2001-03-02
```


Db 1504 -----SISKMSLSYANQEPGILQQKNAVOIISALDTONESTKO-TENTFVLG 1552
QY 1595 QAMEMDLSVLOQVEDLERRVASASLQVKGWMCPEPASERE-----DLVYFEHKSFTKLC 1648
Db 1553 DVQKTD-AFVPVYSDTIQEASPNFE-KAYTLPVLPEKDFNGSDASTQLNTHYAFSKLT 1610
QY 1649 -KEHDGEFTGEDESSAHALERKSDNPLD-IAVTRLADLERNTIERIEEDIAPLRVWRA 1706
Db 1611 YKSSSGHEVENSTTDTQVISHEKENKLESVLTHLS-----RCDSDL----- 1652
QY 1707 LSEARSAQVALCIIQLOQSIaweKSIMKVYQOI---CRKGD-NRELLLLCDGCKGCHT 1762
Db 1653 -----CEMAGMPKGNLNEQDPKHCPSEK----- 1677
QY 1763 YCHRPKITTPDGDWFCACIAKASGQTLKIKKLVKGGKTNFESKKGK-----KVTLTGD 1817
Db 1678 -----CLLSIEDEESQQSILSSLENHSQOSTQPEMHKYGQLVKVELEEN 1721
QY 1818 TEDEDSASTSSSLKRGNDLQKRKMEENTSINLSKQESFTSVKPKRDDSKDLALCSMLL-1877
Db 1722 AEDDKTENQIPQRMTRNK-----ANTMANQSQILASCTLLSEKDSSESSPRGIRL 1773
QY 1878 TE---METHEDAWPELLPVNLKLVPGYKKVKKKMDFTIREKLSGQYPNLETFALDVR 1934
Db 1774 TEDDDPQIHH-----PRKRKVSVPQVQVSPSPQLQAKERTQOSLAAIVDS 1819
QY 1935 LVFDNCETFNEDDDSDIGRAGHNMRYFEKK 1964
Db 1820 LKLEIQPYSSERANPYFYLHIRKKIEK 1849

Search completed: March 9, 2005, 15:09:49
Job time : 199.045 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6

Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 9, 2005, 13:38:21 ; Search time 38.8334 Seconds
(without alignments)
4885.990 Million cell updates/sec

Title: US-10-702-148-21

Perfect score: 10138

Sequence: 1 MGQTKSTSSGGNKKCNQEQ.....AGHNMRKYFEKKWTDFKVS 1972

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 65 summaries

Database : PIR_79.*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2342	23.1	449	T12495	hypothetical prote
2	1090	10.8	1430	T34516	hypothetical prote
3	513	5.1	1479	T17401	transcription regu
4	441	4.3	811	T08738	hypothetical prote
5	378.5	3.7	3259	A56539	giantin - human
6	373	3.7	5327	T13564	microtubule-associ
7	372.5	3.7	3225	I52300	giantin - human
8	358	3.5	1909	A45592	liver stage antige
9	352	3.5	2722	T20532	hypothetical prote
10	351	3.5	2738	E88320	protein F07A11.6 [
11	347.5	3.4	1526	A45605	mature-parasite-in
12	344.5	3.4	3488	T34418	hypothetical prote
13	328	3.2	2526	T20531	hypothetical prote
14	327	3.2	1240	S52734	hypothetical prote
15	326	3.2	3187	J58337	364k Golgi complex
16	314	3.1	1871	D96796	probable heat choc
17	310	3.1	1875	S38173	myosin-like protei
18	310	3.1	2954	T14156	kinesin-related pr
19	309.5	3.1	1898	A45973	trichohyalin - hum
20	309	3.0	2139	T18296	myosin heavy chain
21	306.5	3.0	1378	E88637	protein F53H1.4 [1
22	305.5	3.0	2843	1 RBHUAP	adenomatous polyo
23	301	3.0	2094	S33124	tpi protein - huma
24	300.5	3.0	2663	S28261	centromere protein
25	296.5	2.9	1085	S62516	hypothetical coile
26	295	2.9	2116	A26655	myosin heavy chain
27	294	2.9	1403	T11583	probable translati
28	293	2.9	1818	S73852	hypothetical prote
29	293	2.9	1927	G64585	cag pathogenicity

30	291.5	2.9	1051	2	S55259	TIF1 protein - mou
31	291.5	2.9	1313	2	F96673	hypothetical prote
32	290	2.9	2346	2	T13829	tpi homolog - frui
33	288	2.8	678	2	A54514	glutamic acid-rich
34	287.5	2.8	1690	2	T13030	microtubule bindin
35	287.5	2.8	1974	2	T30010	hypothetical prote
36	286	2.8	4687	1	A39638	plectin - rat
37	285	2.8	1407	1	S28589	trichohyalin - rab
38	285	2.8	1827	2	T16270	hypothetical prote
39	284.5	2.8	1974	2	T16703	hypothetical prote
40	284.5	2.8	3147	2	T18674	hypothetical prote
41	284	2.8	1020	1	QFHUH	neurofilament tripi
42	284	2.8	1359	2	T34036	hypothetical prote
43	284	2.8	2845	2	I49505	adenomatous polyo
44	283.5	2.8	3122	2	T17202	DNA-directed DNA p
45	282.5	2.8	4385	2	T29042	hypothetical prote
46	281.5	2.8	1957	2	T38077	hypothetical coile
47	281.5	2.8	2442	2	T08621	centrosome associa
48	281	2.8	3924	2	S37431	ankyrin 2, neurona
49	280.5	2.8	2253	2	T30336	nuclear/mitotic ap
50	280	2.8	1233	2	T14157	serine/threonine p
51	280	2.8	1804	2	T34518	nestin - golden ha
52	279.5	2.8	1052	1	A44937	kinetoplastr-associ
53	279.5	2.8	1280	2	T00365	hypothetical prote
54	279	2.8	2447	2	T16870	hypothetical prote
55	278.5	2.7	1780	2	S67593	transport protein
56	278	2.7	2103	2	A42184	nuclear mitotic ap
57	276	2.7	1231	2	T18532	serine/threonine pr
58	276	2.7	1684	2	JW0057	gravin - human
59	276	2.7	2774	2	A43359	microtubule-associ
60	275.5	2.7	1642	2	T08880	NMDA receptor-bind
61	275.5	2.7	4574	2	G02520	plectin - human
62	272.5	2.7	1128	2	G86266	hypothetical prote
63	272.5	2.7	1938	2	A59293	skeletal myosin he
64	272	2.7	1805	2	A34736	nestin - rat
65	271.5	2.7	1701	2	T09127	probable erythrocy

ALIGNMENTS

RESULT 1

T12495
hypothetical protein DKFZp434H071.1 - human (fragment)
C:Species: Homo sapiens (man)
C>Date: 23-Jul-1999 #sequence_revision 23-Jul-1999 #text_change 09-Jul-2004
C:Accession: T12495
R:Poustka, A.; Klein, M.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, June 1999
A:Reference number: Z17525
A:Accession: T12495
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-449 <POU>
A:Cross-references: UNIPROT:Q9UIF8; EMBL:AL080173
A:Experimental source: adult testis; clone DKFZp434H071
C:Genetics:
A>Note: DKFZp434H071.1
C:Superfamily: transcription factor GCN5; bromodomain homology
F:366-421/Domain: bromodomain homology <BRO>

Query Match 23.1%; Score 2342; DB 2; Length 449;

Best Local Similarity 99.8%; Pred. No. 2.7e-89;

Matches 448; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1524 RIIDPEDLKALKVHLRGIREKALQKIQKHLDYITQACLNKDKDVAIIELNENEQVT 1583

Db 1 RIIDPEDLKALKVHLRGIREKALQKIQKHLDYITQACLNKDKDVAIIELNENEQVT 60

Qy 1584 RDIENVSVSQANEMDLVLOQVEDLERRVASASLQVKGWMCPEPASEREDLVYFEHS 1643

Db 61 RDIENVSVSQANEMDLVLOQVEDLERRVASASLQVKGWMCPEPASEREDLVYFEHS 120

Qy 1644 FTKLCKEHDGEFTGEDESSAHLERKSDNPLDIAVTRLADLERNIERRIEDIAPGLRVW 1703
Db 121 FTKLCKEHDGEFTGEDESSAHLERKSDNPLDIAVTRLADLERNIERRIEDIAPGLRVW 180
Qy 1704 RRALSEARSAQAVALCTIOQLQKSIWEKSIIMKYVCQICRKGDNBELLLCDDGCKGCHTY 1763
Db 181 RRALSEARSAQAVALCTIOQLQKSIWEKSIIMKYVCQICRKGDNBELLLCDDGCKGCHTY 240
Qy 1764 CHRPIKTIIPDGMFCFACIACASQTLKIKLHVKGKKTINESKGGKVVLTJGTDEBDS 1823
Db 241 CHRPIKTIIPDGMFCFACIACASQTLKIKLHVKGKKTINESKGGKVVLTJGTDEBDS 300
Qy 1824 ASTSSSLKRGKNDLQKRMKMBENTINLSKQBSFTSVKPKRDDSKDLALCSMILTEMETH 1883
Db 301 ASTSSSLKRGKNDLQKRMKMBENTINLSKQBSFTSVKPKRDDSKDLALCSMILTEMETH 360
Qy 1884 EDAPFLLPVNLKLVPGYKVKIKKPMDFSTIREKLSGQYPNLTFFALDVLRLVFDNCETF 1943
Db 361 EDAPFLLPVNLKLVPGYKVKIKKPMDFSTIREKLSGQYPNLTFFALDVLRLVFDNCETF 420
Qy 1944 NEDSDIDGRAGHNRKYPFKKWTDTFKVS 1972
Db 421 NEDSDIDGRAGHNRKYPFKKWTDTFKVS 449

RESULT 2
T34516
hypothetical protein ZK783.4 - Caenorhabditis elegans
C/Species: Caenorhabditis elegans
C/Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 09-Jul-2004
C/Accession: T34516
R/Favella, A.; Vaudin, M.
submitted to the EMBL Data Library, August 1994
A/Description: The sequence of C. elegans cosmid ZK783.
A/Reference number: Z21536
A/Accession: T34516
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-1430 <FAV>
A/Cross-references: UNIPROT:Q23590; EMBL:U13646; PIDN:AAC24421.1; GSPDB:GN00021; CESP:ZK783
A/Experimental source: strain Bristol N2; clone ZK783
C/Genetics:
A/Gene: CESP:ZK783.4
A/Map position: 3
A/Introns: 248/3; 373/3; 547/1; 593/2; 905/2; 1042/3; 1116/3; 1317/2; 1376/1
F:1284-1339/Domain: bromodomain homology <BRO>

Query Match 10.8%; Score 1090; DB 2; Length 1430;
Best Local Similarity 20.6%; Pred. No. 2.6e-37;
Matches 398; Conservative 295; Mismatches 539; Indels 698; Gaps 52;

Qy 147 QKQKQVLSQQLPFIQ---SSQAKE-ESVNHKHTSVIQTGLVSNV--KPLSLVNOAKKE 200
Db 18 QKQKQVLSQQLPFIQ---SSQAKE-ESVNHKHTSVIQTGLVSNV--KPLSLVNOAKKE 200
Qy 201 TYMKLIVSPDVLKAGKNVTSEESLITSELRSREYQKAPPSQLKQSSSKLKVIA 260
Db 78 QIQQQL-----AQGALMAAQ-----KQKQKAAADKAKEKEKOKAAAAA 118
Qy 261 ALSNPKA---TSSSPAHPKOTLENNHPNFTNALLGHQHPNGVIOEAPLALTTKT 317
Db 119 AAAAASASSTSSASALPG-----LSPML-----AAWQQAIOAWALQOMMT 161
Qy 318 KMQSKINENIAAASSTPSSPNVLSSTGRKTPGNTQTPVMPFAS--PILHSQKKEKAVSNV 376
Db 162 POKSOMEAIAIKQMDMAKKPAGVASTSSASTSSSTSTSSASTSSNNNAANNAANMM 221
Qy 377 NPVKTQHSHSPAKSLVQFRGTDSDIPSSKDSNDEDEDEDEDEDEDEDEDEDESDSQ 436
Db 222 NNVMWQ-----LVAQMQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQKQ 262
Qy 437 SEDSNSESDE-----GSEEDDDDDQKQSDSDTEGEKTSMLKNTTSVKSPPMSLT 491

Db 263 KEQDVKNKQOEILKFLMAQHQLNHQKHEKKQADAAA----- 300
Qy 492 GHSTPRNLHTAKPGSAPALCSSESQSPALGTSSTLTSSPHSGTSKRRTVDERELRI 551
Db 301 -----LAAKVLAAHRAAL--ESDSP-----EBGKKTTEAMLRL 331
Qy 552 PLEYGMOETRIIRNFGGR-LQGEVAVYAPCGKKLQYQYEVIKYLSRNGINDISDRNFSFS 610
Db 332 PLQUGWRQTCVRSIASAGVKGDSVPAPCGKKLSTYSEVVRYLTKNHSIHYITRDNFLFN 391
Qy 611 AKIRVGDFY-----EARDGPQEMQWCLLKEEDVIPRAMEGRGRPNPDRQARBEES 665
Db 392 TKLVIGEFIVPKQTEADETQOEREFAMFTEDD----- 423
Qy 666 MRRKGRPPNVGNAEFLDNADAKLLRLQAOETARQAQIKLLRKLQKQEQARVAKK 725
Db 424 -----INKELTRLNLVLFVPK----- 439
Qy 726 QQAIIAAAEKRRKQEQIKIMKQOEKIKRIQOIRMEKELRAQQLLEAKKKKEEAANAKLL 785
Db 440 ---IQASTSGVHEDDIKMSKIEPDEPLDPSELNDEF-----TEELVHQIM 484
Qy 786 E---AEKRIKEKEMRROQAVLLKHQERERRRQHMLMKAMEARKKAEKERLKOEKDEK 842
Db 485 SNGVDECKIRERE-----ADDL 501
Qy 843 RLNKKERLQORLELEMAKELKKPNEDMCLADQKPLPELPRIPLGLVLSGSTFSCLMVQ 902
Db 502 LVN-----INDVRHLPDFSRIGNCQLSSQSGFADALMVHE 535
Qy 903 FLNFGKVLGVDNIDVPNLVLSLQEGELNIGDSNGEVODLLVLLLSAAVCPGLITGYKA 962
Db 536 FVQNFHVLGIDIEI-APKLESICAGLDGDANHAETQLTQLRLRLALEPFGMGNEKRF 594
Qy 963 KTAGLGEHLNVGNRDNVSEILQIFM--EAHCGQTELTSLKTKAFOAHTPAQKASVLAF 1020
Db 595 GQGGGE---MGLDRENFSEVRLFLIDKGRGE-ELSQPLLTNCLFSLISPEQKASILAF 649
Qy 1021 LINELACKSVSEIDKNIDYMNLRDRKVVVEGKLRLKRIIHAHKTGKTDSTGGIDLGE 1080
Db 650 LCDELVCSRNVTETIDKNLDEISRLKGEKWMREGKARALRSARSK--KNDEKVVVVKEE 707
Qy 1081 EQHPLGTPTPRKRRRKGSDYDDDDDDDDQDEDEDEDEDEDEDEDEDEDEDEDEDEDEDE 1133
Db 708 QNH-----ESDSEPTPTPTPKATVAPTIV 734
Qy 1134 -----DICEDEGDOAASVEELEKQIEKLSKQSQYRRKLPDASH 1174
Db 735 SVSPVSAQQQQRKFTPLGQCQEVLTQEESMSLQQQMSLIGDLHQEAQINQKHDTGL 794
Qy 1175 SLRSVMGPDPRYRRYWLPRCGGI FVEGMESEGLEBEIAKEREKLLKAEVQIKEEMFE 1234
Db 795 KIRSFPEGTDFRHRNWMLAHTDKVITESLAT----- 826
Qy 1235 TSGDLNCSNTDCEQKEDLKEDNTNLFQKPGSFSLKLLKLVAKMPPSEVMTPKPN 1294
Db 827 TSVNPNACNAEVA-----SKDPTLEORVPGACETID--LDV----- 862
Qy 1295 AGANGCTLSYQNSGKSHLSGVSQSTATQSNVEKADSNLNTGSSGPGKFFSPLPNDQLLK 1354
Db 863 -----IACVEDLV----- 870
Qy 1355 TLTEKNRQWFSLLPRTPCDDTSLTHADMSTASLVTPOSQPSKSPSTPAPLPGSSAQNVP 1414
Db 871 -----DDVLLRA----- 878
Qy 1415 GLNPFPALPLQVKGVSMMGLQFCGWPTGVVTSNIPTLSVPSLGLSGLSEGNNGSFLT 1474
Db 879 ----- 878
Qy 1475 SNVASSKSESPVQNEKATSAQPAAVEKAVPDVFPSPKPIPEEMQFGWRIIDPEDIKAL 1534
Db 879 -----KADKTKRKYRRIE-----NHMKRGWWTMQNRDCVESL 911


```
Qy 1360 NRQFSLRLPTPCDDTSLTHADMSTASLVTQSQPPSKSPPTAPLGSSAQNVPVGLNPF 1419
Db 892 -----IAKALVLRRT-----PIGTDRNH-----NRY 913
Qy 1420 ALSPLQVKGVSMMGLQFCGMPGTGVTNSIPTLSVPSLGLGLSEG---NGNSFLTNS 1476
Db 914 WL-----FSNEVP-----GLFIEKGWVHNSIDYRFKH 940
Qy 1477 VASSKSRSP-----VPONEKATSAQAAVEVAK-PVDFPSPKPIPEEMQFGWRRIIDPEDL 1531
Db 941 HRKDHSLPDDDDYCFRRKCANLGNASVNAHGGPALEAVETTVPKQGNLWFLCDSQKEL 1000
Qy 1532 KALLKVLHLRGIGREKALQKQIQKHLDYITQA--CLKNKDVAIIELNENE--NOVTRDIV 1507
Db 1001 DELSLCHLPQGISQFKELEKRYQELTHCNMARKPNLGLKSCDGNQELNLFRLSDLI 1060
Qy 1588 ENWSVEEQAMEMDLVLQOVEDLERRVAS-----ASLQ-----VKGMWCPPEP 1629
Db 1061 E---VATRLQKGLGYMEGTSEFEARVLSLEKLDGFCGVIALQASVIKKFLQGFMAP-- 1115
Qy 1630 ASEREDLVYFEHFKFTKLCKEHDEGFTGEDSSAHLERKSDNPLDIAVTRLADLERNIE 1689
Db 1116 -----KOKRKLQSEDSSTKSEEVDEE-----K 1137
Qy 1690 RRIEE-DIAPGLRVRRALSARSAAQVALCQLOKLSIAWEKSIMKVYCOICRKGDNNEE 1748
Db 1138 KQVBEAKVASALEKWKTAIREAQTFSRHVLGLMDACIKWDMAENARCKVCRRKGEDD 1197
Qy 1749 LLLLCDCGDCGCHTYCHRPKITTIPDGDWFCPACIAKASGQTLK----IKKLHVKGKKTN 1804
Db 1198 KLILDCENKAFHLFCLRPALYEPDGEWQCPACQPTARRNSGRNYTEESTSEGSGD 1257
Qy 1805 ESKGKKVTLTGDTEDDSASTSSLK-----RGNKDL-----QKRKM 1842
Db 1258 ESGEEBEEBEEBEEDYEAGLRLPRKTRIGKQSVIPAARPGRPKGKSHPARSRP 1317
Qy 1843 BENTSINLSKQESTSVKPKRDSKDLALCSMLTETHEDEDAMPFLLPVNLKLVPGYK 1902
Db 1318 KQDEPVD-----DLVLTQTRISRRSLOKEDILHLKLVYRFSVPREPVRDRDEADYY 1373
Qy 1903 KVIKKPMDFTIREKLSSGOYPNLETFALDVLRLVDFNCETFN 1944
Db 1374 DVIEHPMDFTIQNKSCGNYSRVSQEFITDMQVFNAAELYN 1415
```

RESULT 4

```
T08738
hypothetical protein DKFZp586E0518.1 - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 09-Jul-2004
C:Accession: T08738
R:Wambutt, R.; Heubner, D.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, March 1999
A:Reference number: 216471
A:Accession: T08738
A:Molecule type: mRNA
A:Residues: 1-811 <WAM>
A:Cross-references: UNIPROT:Q9NRL2; EMBL:AL050089
A:Experimental source: adult uterus; clone DKFZp586E0518
C:Genetics:
A:Note: DKFZp586E0518.1
F:709-764/Domain: bromodomain homology <BRO>
```

```
Query Match 4.3%; Score 441; DB 2; Length 811;
Best Local Similarity 20.2%; Pred. No. 5.7e-11;
Matches 202; Conservative 108; Mismatches 270; Indels 418; Gaps 31;
Qy 1138 DEDGDAQAAVEELEKQIEKLSKQSQVRRKLFPAASHLSRSLVMPGPRYRRRYWILPRCG 1197
Db 28 DEEB---ALKQEHQKKELEKIQS-----AIACNIFPLGRDRMYRYYWIFPSIP 76
Qy 1198 GIFVEGMESGBGLEIAKEREKLAESVQIKEEMFET-SGDSLNCSTNDHCEQKEDLKE 1256
```

```
Db 77 GLFIE--EDYSGLTEDMLPRPSSFQNVQSQDPQVSTKTGEPLMSBESTNIDQG----- 129
Qy 1257 KDNINLFLQPGSFSLKLELVAKMPPSESVMTPRNAGANGCTLSYQNSGRKSLGSVQ 1316
Db 130 -----PRDHSVLQPKP----- 140
Qy 1317 STATQSNVEKADSNLNTGSSGPK--FYSPLPN-DOLLKTLTEKUNOFSLPRTPPCD 1373
Db 141 -----VHK-----PNRWCFFSSCEQLDQLEALNSRGRHSAL-----K 174
Qy 1374 DTSILTHADMSTASLVTQSQPPSKSPPTAPLGSSAQNVPVGLNPFALSPLOVKGVSMM 1433
Db 175 ETLLQESRICAQLARFSEBEKFHFDKQPD----- 205
Qy 1434 GLOFCGMPGTGVTNSIPTLSVPSLGLGLSEGNNSFLTNSVASSKSPVPONEKAT 1493
Db 206 -----SKPTYSRG-----RSSNAYDPSQMAEKQ----- 229
Qy 1494 SAQPAAVEVAKPVPDFPSPKPIPEEMQFGWRRIIDPED-----LKALLKVLHLRGIREKA 1547
Db 230 -----DEL-RLRDF-----LQIEDRIYQGTIGAL----- 253
Qy 1548 LQKIQKHLDYITQACLKNKDVAIIELNENEENQVTRDIVENMSVEEQAMEMDLVLQOV 1607
Db 254 --KVTDRH---IWRSALESGRYELLS--EENKENGIIKTV--NEDVEEMEIDEQTKVIVKD 305
Qy 1608 EDLERRVASASLQVKGMCPPEASEREDLVYFEHFKFTKLCKEHDEGFTGEDSSAHL 1667
Db 306 RLLGKITETPSTVSTNASTPQSVS-----SVWHYL-- 335
Qy 1668 RKSDNPLDIAVTRLADLERNIERRIEDIAP-----GLRVRRALSARSAAQ 1715
Db 336 -----AMALFQEQGIERFLK--APLDASDSGRSYKTVLDRWRRESLSSASLSQ 383
Qy 1716 VALCQLOKLSIAWEKSIMKVYCOICRKGDNBEILLCDGDCGCHTYCHRPKITTIPDG 1775
Db 384 VFLHSLFLDRSVIWSKILNARCKICRKGDAENWVLCDGCDRGHHTYCYRPKLKTVP 443
Qy 1776 DWFCPACIAKASGQTLKIKKLH--VKGKTNESKKGKVTLTGDT-----D 1820
Db 444 DWFCPCRPQRSLRSLRQPSLESDVEDSMGDEDDVDGDEERGQSEEEYEVEQD 503
Qy 1821 ED-----S 1823
Db 504 EDDSQSEEEVSLPKRGRPVRLPVKTRGKLSSESFSRSGQQQEPGRYPSRQQSTPKTTVS 563
Qy 1824 ASTSSSL-----RGNKDLQK-- 1839
Db 564 SKTGRSLRKINSAPPTETKSLRIASRSTRHSHGHPLOADVVELLSPPRRKGRKRSANKTP 623
Qy 1840 -----RKMEENTSIN-----LSQESFTSVKPKRDKDSK----- 1868
Db 624 ENSPNFENFRVIATKSEQSRSVNIASKLQSESKRCKRCKQSPSPVTLGRRSSGR 683
Qy 1869 -----DLALCSMLTETHEDEDAMPFLLPVNLKLVPGYKVKVKKPMDFTIREKLSSGOY 1923
Db 684 QGGVHELSAFEQLVVELVRHDDSWPFLKLVSKIQVPDYDIILKKPIALNIREKVNKEY 743
Qy 1924 PNLETFALDVLRLVDFNCETFNEDSDIGRAGHNNRKYF 1961
Db 744 KLASEFIDDIELMFNSCNCFEYNPNTSEAKAGTRLQAF 781
```

RESULT 5

```
A56539
giantin - human
N:Alternate names: macrogolin
C:Species: Homo sapiens (man)
C:Date: 19-Oct-1995 #sequence_revision 26-Jan-1996 #text_change 09-Jul-2004
C:Accession: A56539; S37536
R:Seelig, H.P.; Schranz, P.; Schroeter, H.; Wiemann, C.; Griffiths, G.; Renz, M.
Mol. Cell. Biol. 14, 2564-2576, 1994
```


1757 QSIPTGETEEOQSLMS---TRPTCSVPSAKSANPAVSKDFSSHSHDEINNYLQIIDQLKE 1813
1060 RIHAKXTGKEDTSGGIDLGEQHPGLTPTTGRKRRKGGSDYDDDDDDSDQGDDEDD 1119
1814 RIAGLEBEKQNKKEFSQTLNEKNTLLS-----QISTKOGELKMLQBEVTTQNNLLNQIQ 1868
1120 ED-----BEDKEDQKGGKTDICEDBEDGQAQAEVEELEKQIEKLSKQSQYRRK 1168
1869 EELSRVTKLKTAEAEKDDLE-----ERLNNQLAELNGSIGNYCQD 1909
1169 LFDA---SHSLRSVMFGPDYRRRYWILPRCGGIFVFGMESGEGLEBIAKEREKLKKAES 1225
1910 VTDQAIKNELLESEMKN-----LKKC---VSELE--EEKQLVKEKTKV---ES 1950
1226 VOIKEEMFEISGDSLNCSTNTHC--EQEKDILKEKNTNLFLQKP-----GSFSKLSKL 1276
1951 EIRKEYLEKTOGAQKEFGNKSHAKLOELLKBEQEVKQLOKOCIRYOEKISALERTVKA 2010
1277 LEVAKMPPESVMTPKPNAGANGCTLSYQNSGKHSLSG--VQSTATQSNVEKADSNLLEN 1334
2011 LEFVQTESQKDLITKENLAQ--AVEHKKAQAELASFKVLLDDTQSEARVLADNL-- 2065
1335 TGSSEGPKFYSPLNDQLLTKLTENKQWFSLLPRTPCDTSLTHADMSTASLVTPOSQP 1394
2066 -----KLKKELQSNKESVK-----SQMKQKDEDLERRLEQAEEK 2099
1395 PSKSPSTPAPLGSSAQNPVCLNFPALSPLOV---KGVSNMGLQFCGWPVTGVTSNIPP 1451
2100 HLKKEKNMQEKLDAIRREKHVLEB--TIGETQVTLNKXDKVEQQLO----- 2143
1452 TLSVPSLGSGLGSEGNGLSPLTNVASSKSESVPQNERKATSAQPAAEVAKPVPFPSP 1511
2144 -----ENLDSVTQLAAPTMSMSSLDQDRD----- 2168
1512 KPIPEMQFGWRIIDPEDLKALKVLHLRGIREKALQKQIQKHLDYITQACLKNKDVAI 1571
2169 -----RVID-----EAKWKERKPSD-----AIQSKEE-- 2190
1572 IELNENEEN--QVTDIVNHSVERQAMEMDLSVLQOVEDLERRVASASLVKQHMGC----- 1626
2191 -EIRLKEDNSVLKQDLQKQMSIHMEELKINISRLHDKQIWESKAQTEVOLQKQKVCDTLQ 2249
1627 ---PEPASEREDLVYFEHKSFTKL-----CKEHDGSEFTG--- 1657
2250 GENKELLSQLEETHRLYHSSQNEVLAKLESELKSLKDQLTDLNSLEKCKQKQKNGLEIIR 2309
1658 ----EDESSAHALERKSDNPLDIAVTRLADLE--RNTERRIEEDIAPLGRVWRALSEAR 1711
2310 QOEADIQNSKPSYBQLE-----TDLQASRELTSLRHEBINMKQEKIISLLSGKE 2358
1712 SAAQVABC--IQLOKSIANEKSTMKVYQCICRKGDNBEILLLLCDGCDKGCHTYCHRPKI 1769
2359 EATQVATAELRQQHDKSI---KELENLLSQ-----EEENIVLEEENK----- 2399
1770 TTIPDGDFWFCPACTAKASG--OTLK--IKKLHVKGK-----KTNESKKGKGVTLTGDT 1819
2400 -----AVDKTNQLMETILTKTKENIQQAQQLDFVKSMSLQNDRDRIQVGYQ 2447
1820 D-----EDSASTSSSLK-----RG----- 1833
2448 QLEERHLSIILEKQDLIQEAENNNKLKEEIRGLRSHMDLNSNAXLDAELIQYREDLN 2507
1834 -----NKDLOKR-----KMEENTSINLSKQSFSTSVKKPKEDDS 1867
2508 QVITIKDSQQQLLEVQLQKNELENYAKLEELKKESEANEDLRRSFNALQBEKQDLS 2567
1868 KDLALCSMILTME 1880
2568 KETESLKVSIISQL 2580

RESULT 8
A45592

liver stage antigen LSA-1 - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:Date: 22-Nov-1993 #sequence revision 02-Dec-1994 #text change 09-Jul-2004
C:Accession: S24597; A45592; S29393; S34842; B45592; C45592; D45592
R:Zhu, J.; Hollingdale, M.
submitted to the EMBL Data Library, November 1990
A:Reference number: S24597
A:Accession: S24597
A:Molecule type: DNA
A:Residues: 1-1909 <ZHU>
A:Cross-references: UNIPROT:Q25893; EMBL:X5G203; NID:G9915; PID:G9916
R:Zhu, J.; Hollingdale, M.R.
Mol. Biochem. Parasitol. 48, 223-226, 1991
A:Title: Structure of Plasmodium falciparum liver stage antigen-1.
A:Reference number: A45592; MUID:92107224; PMID:1840628
A:Accession: A45592
A:Molecule type: DNA
A:Residues: 1-195;638-688;1165-1215;1590-1909 <ZH2>
A:Note: sequence extracted from NCBI backbone (NCBIN:75010, NCBIN:75012, NCBIN:75014, NC
R:Guerin-Marchand, C.; Drulhe, P.; Galey, B.; Londono, A.; Patarapotikul, J.; Beaudoin,
Nature 329, 164-167, 1987
A:Title: A liver-stage-specific antigen of Plasmodium falciparum characterized by gene c
A:Reference number: S29393; MUID:87315391; PMID:3306406
A:Accession: S29393
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 323-387 <GUE1>
A:Cross-references: EMBL:M28266
R:Guerin-Marchand, C.; Drulhe, P.; Galey, B.; Londono, A.; Patarapotikul, J.; Beaudoin,
submitted to the EMBL Data Library, April 1992
A:Description: a liver-stage-sepcific antigen of plasmodium falciparum characterized by
A:Reference number: S34842
A:Accession: S34842
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 323-381, 'HKAI', <GUE2>
A:Cross-references: EMBL:M28266
A:Note: difference at carboxyl end due to frameshift error
C:Comment: This protein is found as flocculent material in the parasitophorous vacuole.
C:Superfamily: trichohyalin; calmodulin repeat homolog
C:Keywords: EF hand
F:154-1629/Region: 17-residue repeats (A-K-E-K-L-Q-E-Q-Q-S-D-L-E-Q-E-R-R)

Query Match 3.58; Score 358; DB 2; Length 1909;
Best Local Similarity 18.3%; Pred. No. 4e-07;
Matches 389; Conservative 345; Mismatches 839; Indels 548; Gaps 76;

Qy 15 KCNQSKNQPLDARVDKIKDKKP-----RKKAMESSNSDSGTSDDTSEGIS 66
Db 113 KENKLNKEGLIEHIINDDDDKKYIKGQDENQEDLEEKAAKETLQGGQSDLEQERLAK 172
Qy 67 SDSDLEEDDEEDQSTIESEDDSDSESAQKSNQVLLHGISDPKADGQKATEK--- 123
Db 173 EKLQEQSDSEQLAKEKLQEQSDLEQERLAKELQEQSDLEQERLAKELQEQSD 232
Qy 124 -AQEKRIHQPLAPESQTHSFQSQKQPQVLSQLPFIQSSQAKESVNNKHTSVLQST 182
Db 233 LEQERRAKEKL---QEQQSDLEQERRAKEKLQEQSDLEQERRAKEKLQEQSDLEQ 288
Qy 183 GLVSNVPLSLVNAQAKETVMKLVPSDPLKAGNKNTSESSLLTSELRSKRQYKQAF 242
Db 289 LAXEKLQEQSDLEQERRAKEKLQEQSDLEQERRAKEKLQEQSDLEQER 339
Qy 243 PSOLKQKQESSKSLKKVIAA---LSNPKATSSSPAHPKQTLENNHPNPLTNALLGNHPN 299
Db 340 LAXEKLQEQSDLEQERRAKEKLQEQSDLEQERRAKEKLQEQSDLEQER 382
Qy 300 GVLTQSVIQEAPLALTTKTKMQSKI--NENIAAASSTPSSPVNLSTSGRRTPGNTQPVMS 358
Db 383 ---QSDLEQERLAKELQEQSDLEQERRAKEKLQEQSDLEQERRAKEKLQEQSDLEQ 439
Qy 359 ---ASPTLHSQ-----GKEKAVSNVNNVPVKTOHHSHHPAKSLVQFRTGTDSDIPSSK 406

Db 440 ERLAKEKLEQOQSDLEQERRAKEKL-----EQOQSDLEQERRAKEKLEQOQSDLEQER 492
Qy 407 DSEDNEDEDEED-----EDDDDDSDSQSDSNSESDTEGSE-----EED 454
Db 493 LAKEKLEQOQSDLEQERRAKEKLQOQSDSEQLERAKEKLEQOQSDLEQERRAKEKLEQ 552
Qy 455 DDDKDD-----ESDTEGKTSM-KLNTTSSVYKSPMS-----LTHST-----PRNL 499
Db 553 QSDLEQERRAKEKLEQOQSDLEQERRAKEKLEQOQSDLEQERRAKEKLEQOQSDLEQERRL 612
Qy 500 HIAKAPG-----SAPALCESQSPAFLTSSSTLTSSPSHGTSKRRV 543
Db 613 AKELQOQSDLEQERRAKEKLEQOQSDLEQERRAKEKLEQOQSDLEQERRAKEKLEQOQSDLEQ 672
Qy 544 TDERELRIPLE-----YGMQRETRIRNFGRLQGEVAYAPCGKKLQYPIVILKLSRN 597
Db 673 SDLEQERRAKEKLEQOQSDLEQERRAKEKLEQOQSDLEQERRAKEKLEQOQSDLEQERRAKE 724
Qy 598 GIMDISRDNFSAKIRVGDYEARDQOEWQWCLLKEEDVIPRAMEGRGRPPNDPR 657
Db 725 ---DLEQER-----RAKELQOQSDLEQ-----BRAKEKLEQ 756
Qy 658 QRAREESMRKGRPPNVGNAEFLDNADAKLRLKLAQEIARQAQIKLLRKLQO-----714
Db 757 QSDLEQERRAKEKLEQOQSDLEQERRAKEKLEQOQSDLEQERRAKEKLEQOQSDLEQERRAKE 793
Qy 715 -EQARVAKE-AKKQQAIIAABEKKQK--EQIKMKQO-----EKIKRIQIIRMEKELRAQ 766
Db 794 LEQDLAKEKLEQOQSDLEQERRAKEKLEQOQSDLEQERRAKEKLEQOQSDLEQERRAKE 852
Qy 767 QILE-----AKKKKEEANAANAKLLBAEKRIKEKEMRQOAVLLKHQERRR---812
Db 853 EKLOEQOQSDLEQERRAKEKLEQOQSDLEQERRAKEKLEQOQSDLEQERRAKEKLEQOQSDLE 904
Qy 813 ---RQMLMAMAKARKEK--LQEKDEKELANKER--LEQRLELEWAKELKPN 867
Db 905 KLEQOQSDLEQERRAKEKLEQOQSDLEQERRAKEKLEQOQSDLEQERRAKEKLEQOQSDLE 964
Qy 868 EDMCLADQKPLPPLPRIPGLVLSGTSFSDCLMVQVFLNFGKVLGFDVNIQVNLVLQ 927
Db 965 EQERRAKEK-----LQ 976
Qy 928 GLNIGDSMGVQDILLVRLLSAUCDPGLITGYKAKTALGEHLNVGNRNVNVEILOIF 987
Db 977 -----QOQSDLEQERL-----AKEKLEQOQSDLEQER--LAKEKLEQ--1009
Qy 988 MEAHGOTELTSLTKAFOAHTPAQKASVLAFLINELACSKSVVSEIDKNIDYNSLR 1047
Db 1010 -----QOQSDLE-----QERRAKEKLEQOQSDLEQERRAKEKLEQOQSDLEQERRAKE 1057
Qy 1048 DKWVEGKRLRIIHAKTGKRTSGGIDLGEEHPLGTPTPGKRRKRGSDSDYDDDD 1107
Db 1058 KLEQOQSDLEQERRAKEKLEQOQSD-----DLEQERL-----AKEKLEQOQSDLEQER 1104
Qy 1108 DDDSDQOQDEDEE--DKSDQKKTIDI CED-----EDEDQAASVEELKEQIKLSQ 1161
Db 1105 LAKEKLEQOQSDLEQERRAKEKLEQOQSDLEQERRAKEKLEQOQSDLEQERRAKEKLEQ 1164
Qy 1162 QS---QYRKLFDASHLSRVNFGPDYRRYRRIWILPRCCGIFVWGESGEGLERAKERE 1218
Db 1165 QSDLEQERR-----AKEKLEQOQSDLEQERRAKEKLEQOQSDLEQERRAKE 1210
Qy 1219 KLKKAES-----VOIKEMFTSGD--SLNCSNTDHCQKEDLKEKNTNLFLOKPGSFS 1271
Db 1211 KLEQOQSDLEQERRAKEKLEQOQSDLEQERRAKEKLEQOQSDLEQERRAKEKLEQOQSDLE 1270
Qy 1272 KLSKLELVAKMPPESEVMTPKPNAGCTTLYSQNGHSHLSGVQSTQATQSNVEKADSN 1331
Db 1271 EQERRAKEKLEQOQSD-----LEQERRAKEKLEQOQSDLEQERRAKEKLEQ 1315
Qy 1332 LFNTGSSGFGFYSPLPNDQLKLTLEKRWQFSL---PRTPCDDTSLTHADMSTASLVT 1389
Db 1316 -----EQOQSDLEQERRAKEKLEQOQSDLEQERRAKEKLEQOQSDLEQERRAKE 1362

Qy 1390 PQSQPPSKSPPTPAPLGSSAQNVPGLNPPFALSPLQVKGVSMMGLQFCGMPGTGVVTSNI 1449
Db 1363 EKLQ-----EQOQSDLEQERRAKEKLEQ-----1385
Qy 1450 PFTLSVPSGLSGLSEGNNGNSFLTNSVASSKSPVPQNEKATSAQPAAVEVAKPVDPP 1509
Db 1386 -----QSDLEQDLAKEKLEQOQSDLEQERRAKEKLEQ-----QOQSDLE 1424
Qy 1510 SPKPIBEMQFGWRIIDPEDLKALLVHLRGRERAKLQIQKHLHYITQACLKNDV 1569
Db 1425 QERRAKEKLEQOQSDLEQERRAKEKLEQOQSDLEQERRAKEKLEQOQSDLEQERRAKEKL 1484
Qy 1570 AIIELNENENQVTRDIVENMSVEQAMEMDLSVLQOVEDLER-RVASASLQVKGWMCPE 1628
Db 1485 QEORQDLQERRAKEKLEQOQSDLEQERRAKEKLEQOQSDLEQERRAKEKLEQ-----1536
Qy 1629 PASREDLVYPEHKSFTYKCKEHGDEFTGEDESSAHALERKSDNPLDIATVTLA-----1682
Db 1537 --EQOQDL-----EQERRAKEKLEQOQSDLEQERRAKEKLEQOQSDLEQERRAKEKLE 1588
Qy 1683 ---DLERNIERRIEEDTAPGLRVWRALSEARSAAQVALCIQOQKSTAWEKSIMKVYCO 1739
Db 1589 QOQSDLEQ--ERLAKK-----LQOQSDLEQERRAKEKLEQOQSDLEQERRAKEKLE 1639
Qy 1740 ICRKGDNEELL-----LCDGCDKGCTYCHRPKITTTIPDGDWFCPACIACA 1786
Db 1640 --RKKEHGVLAEDYGRLEIPALPELPSENERGYI-----PHQSSLPQDN-----1683
Qy 1787 SGQTLKIKLHVKKKTNES-----KKGKVTLTGDTEDDEDSASTSSSLKRG-NKDLOKRG 1841
Db 1684 RGNRSKSEISIIKTNRESITTNVEGRDTHKGLHEKKGDSIKPEQEKDSADIQNHT 1743
Qy 1842 MEENTSINLSQESFTSVK-----KPKRDDSQKALCSMILTEMETHEWPFLLPVNLK 1896
Db 1744 LE---TNISVNDVQISKYEDEISAAYDSS-----LDEDEDEDDLEF-----1785
Qy 1897 LVPQYKVIKPMDFSTIREKLSGGQYPNLFTALDVLRFVDFNCETFNEDSDSIGHAGHN 1956
Db 1786 -----KPIQVQNFQDEENIGYKELED-----LIEKNENLDDLEGEIKSSEE 1829
Qy 1957 MR-----KYFEKKTDTDFK 1970
Db 1830 LSEKIKKKGKYEKTDKNFK 1850
RESULT 9
T20532
hypothetical protein F07A11.6b - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T20532; T27777
R:Palmer, S.
submitted to the EMBL Data Library, October 1995
A:Reference number: Z19287
A:Accession: T20532
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2722 <W1>
A:Cross-references: UNIPROT:Q19135; EMBL:Z66511; PIDN:CAB54211.1; GSPDB:GN000020; CESP:F07A11.6b
A:Experimental source: clone F07A11
R:Gajadaty, S.
submitted to the EMBL Data Library, March 1996
A:Reference number: Z20417
A:Accession: T27777
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2722 <W12>
A:Cross-references: EMBL:Z69904; PIDN:CAB54502.1; GSPDB:GN000020; CESP:F07A11.6b
A:Experimental source: clone ZK20
C:Genetics:
A:Gene: CESP:F07A11.6b
A:Map position: 2

A;Introns: 36/2; 92/2; 182/2; 272/3; 344/2; 426/2; 490/1; 541/2; 700/3; 770/3; 1286/3; 1	
Query Match 3.5%; Score 352; DB 2; Length 2722;	
Best Local Similarity 18.6%; Pred. No. 1.1e-06;	
Matches 349; Conservative 235; Mismatches 642; Indels 648; Gaps 72;	
Qy 4	TKSTSSGGNKRKNQEOSKN--QPLDARVD-----KIKDKKPRKAMESSNSDSGTS 56
Db 383	SKSENDQSSSPSSRRQNLHDLPTQSSVEHHTNQEDQENNASGSDSSDSSEBSS 442
Qy 57	SDTSSGSSSDDLDEDEEEDQSEEDSDSEBAQHKNQVLLHGIDSPKAD 116
Db 443	S-----SNBDSQNDVDEDDVYSEKHEPEBEGKSSPGN-----GHRD-ESN 488
Qy 117	GOKATEAQEKRIHQLPLAFESQTHS-----FQSQKQPOVLSQLP-FIQSSQAK 168
Db 489	GDKDHESSE-RFSQPTSSHHETSHSPEKDSEAYQSRFSPLNYQSQSPGYELESKEI 547
Qy 169	EESVNHKTSVIQSTGL-----VSNVPLSLVNOAKK 199
Db 548	KQESPTSSASSDLELDMEPDNLPLRLERHWRPFIDVSSFVRIDEIVELNOKAR 607
Qy 200	ETYNKLI-VSPDVLKAGNKVNTSESLTSELRSREYQKAPP-SQLKKQESSKSLKK 257
Db 608	ASYEKFTGRPPPKC-----NNDEVLSTQKIVFHEPRDYVYENPCSELEVRIDWRKLS 661
Qy 258	VIALSNPKATSSPAHPKQTLNHNHPNPLTNALLGNHQPNGVIOVQEAPLATTKT 317
Db 662	DTADLDDFRATDSKE-----LGRDQAGGRTSGRPSLDESRTNRL 701
Qy 318	KMQSKIN-ENIAAAS-----STPP-SSPVNLSTS---GERTP-----GNQTPMP 357
Db 702	SFDTHTHPAELAQRSLSLCIGPWTPTPTPTSPQPLVNTLHPQTSPSTSGGTTTPRS 761
Qy 358	SASPILHSQKQKAVSNVNPVKTHSHHPAKSLVEQFRGTDSDIPSSKDSSENEDEE 417
Db 762	QPPPLMSPVSRHNSMSTGRPASITQLRHQSVMFPDVPDVSIPPPPIPTHDEMA- 815
Qy 418	DDEEDEDDEDDSDSDSESNSDSTEGSEEDDDKQDQSDSDTEGETSKMLN 477
Db 816	-----PRGTPPSRRSS 826
Qy 478	KTTSVSKSPMSLTGHTSPRLNLIHAKAPGAPAAALCSQSAPFLGTSSSTLTSSPHSGT 537
Db 827	ETWVPLASPP----GTPIQ-NULLTWPVIPPPLI-----AATSTGTHSVSSAHS-- 872
Qy 538	SKRRVTDRELRIPLYGWQRETRIRNFGRLQGVAYVAPCGKLRQYPEVIKYLARN 597
Db 873	TPRHSIS-----GTPVHCEPSNSKTSQPPTP---KSRP 902
Qy 598	GIMDISRDNFSFSAKIRVGDFYEARDGQPOEMQWCLLKEEDVIPRIAMEGRGRPPNPDR 657
Db 903	EKVQIRHDTIS-----KSGPSN-----AINALQARSQMSGTSDP 936
Qy 658	QRARESRMRKRG-----RPPNVGNAEFLDNADAKLLRLQAOETARAOAOKLL 708
Db 937	KKSAPSTPVVDAGSDLVIAQIMSQPLG-----LRKL--PRIEKSSALQNI 982
Qy 709	RKLQOBQARVAKBAKQQAIAAEKRRKQKEQIKIMKQOEKIKRIQQIIRMEKELRAQOI 768
Db 983	QNHQPPHSNANSTPSTPSTSTHQAMFKDKKERKK--KEKEKEEREREAREMK- 1034
Qy 769	LEAKKKKKEEANAALLAEKRIKEKEMRQQAALLKHQERRRRQHMLMKAMEARKKA 828
Db 1035	---RKETEERNKRKEMERAKRLEDERQERKRE---KKERDERK-----KEKEKVRKK 1082
Qy 829	EKEERLQOEK-----ROEKLNRKLEKQRELEMAKELKPNEDMCLADQKPLPE 880
Db 1083	AEKELKKKKHRRKGDSDSDSDSDNDELDDVRSKTEMTOEEK-----DHQLA- 1131
Qy 881	LPRIPGLVLSGSTFSDCLMVVQFLRNFGKVLGFDVNDIVPNLSVLQEGLLNIGDSMGEVQ 940
Db 1132	-----LLLSKGGI-----IENLKSRRRSKRAHDSPEKMQ 1161

Qy 941	DLIV--RLLSAAVCDPCLITGYKAKTALGEHLLNVGNRDNVSEILQIFMEAHCGQTILT 998
Db 1162	QKSQORVLLESSDDEGKDGDKGNSNGE----- 1191
Qy 999	ESLTKAFOAHTPAQKASVLAFLINELACSKSVVSEIDKNIDYMSNLRRDKWVVEGLRK 1058
Db 1192	ESDSEKADLPPPPAPPS-----LSASADQR 1216
Qy 1059	LRIHAKTKKRDTSGGIDLGESQHPGTPTGPKRRRKGSDSDYDDDDDDSDQDQ-E 1117
Db 1217	LKVLKEREKGELTTS-----SDDEHNDAGEIH 1244
Qy 1118	DDEDEKEDQKGGK--TDICEDE-----DEGQQAASVEE--LEKQIEKLSK 1160
Db 1245	QORLTERENRKKQKSLTAYSSDEQGERKNVPKMRRDDSEAAKHPGWSAKDDQKQRK 1304
Qy 1161	QSQYRKLFDASHLSRSMVFGPDYRRRYWILPRCGGI FVEGMESGEGLAEIAKEREKL 1220
Db 1305	RKLEHRSSSEDESK-----NAKDRFRDIPH-----EDVSDEETE 1340
Qy 1221	KKAESVOIKEEMPETSGDSLNCNTDHCROKEDLKEKDNLTNLFLOKPGSFKLS-KULEV 1279
Db 1341	DGSRSS-----RRQSTSTISNVTAKEKKE--KSGKTLPLRIVPEPTGTPLSPKILSP 1390
Qy 1280	AKMPES-----EYVTPKNAGANGCTLSYQNSGKHSGLSVOS-----TA 1319
Db 1391	KHLSPTKSTSTSTRKSSISDHENLISPRQRNTSTSTATTSSKHEALSIEPEKLSPPVT 1450
Qy 1320	TQSNVEKADNNL---FNTGSSGPGKEYSPLPNDQLLTKLTKTEKNRQWFSLLPRTPCDDTS 1376
Db 1451	AKSSVSIIDPISRDEFSMNSAAD---SPMSTTGRPMVLTKAAMKAFNSTPPKKKNSS 1506
Qy 1377	LTHADMTASLVTQSQPPSKSPSPAPLPGSSAQNPVGLNPFPALSPLOVKGVSMMGLQ 1436
Db 1507	GQH-DSSSGS-----SSDSSSDGSTSSDSDDEVPKQTEPVTSPVVASDN----- 1553
Qy 1437	FCGWPTGVVTSNIPFTLSVPSGLSGLSGEGNSFLTSNVASSKSPVPONEKATSAQ 1496
Db 1554	--GSPENV-----VETFSI---VSQTPREPPFTTISQSSSESEPEAVPECEPA-SVE 1600
Qy 1497	PAVEVAKPYDFPSPKPIPEMOFGWMWRIIDPEDLKALLKVLHLRGIREKALQKIQKHL 1556
Db 1601	P-QMETSQNV-----PVSEE-----HED-----SHEHGDSEAVESQQQPLE 1637
Qy 1557	DYTOACLQNK--DVAIIELNEENQVTRDIVENSVSEQAMEMDLVLQOVEDLERRV 1614
Db 1638	HOBEKELENKILDVAA---EHHEEQVQGD-----EDSVESI----- 1672
Qy 1615	ASASLVKGMWCMCEPASEREDIVFHKSTKLCKE----- 1650
Db 1673	-----PAPSEDPVTTQAEKSAHTLISDQETQAVQSIFDEEAEFPQVPP 1720
Qy 1651	-----HDGEFTGBD-----ESSAHLERKSDNPLDIAVTRLADL 1694
Db 1721	FGISTNEKEVSGKDPHNKPTPLNNGHTDLLFSPSSSAHASEKQTSKSD-----DM 1773
Qy 1685	ERNIERRIEADIAP 1698
Db 1774	EEDSELVVMKEVP 1787

RESULT 10
E88320
protein F07A11.6 [imported] - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 10-May-2001
R;Accession: E88320
R;anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A;Title: Genome sequence of the nematode C. elegans: a platform for investigating biology
A;Reference number: A75000; PMID:99069613; PMID:9851916
A;Note: see websites genome.wustl.edu/gsc/C_elegans/ and www_sanger.ac.uk/Projects/C_elegans/

A>Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
A:Accession: E88320
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2738 <STO>
A:Cross-references: GB:chr_11; PIDN:CAA93781.1; PID:g3881547; GSPDB:GN00020; CESP:F07A11
C:Genetic8:
A:Gene: F07A11.6
A:Map position: 2

Query Match		3.5%;	Score 351;	DB 2;	Length 2738;
Best Local Similarity		18.7%;	Pred. No. 1.2e-06;		
Matches 356;		Conservative 227;	Mismatches 627;	Indels 696;	Gaps 72;
Qy	4	TKSTSSGGNRKCNQEQSKN	---QPLDARD---	KIKDKPRKAMESSNSDS	SGTS 56
Db	383	SKSENDQSSSPSSRRQNLHDLQTRSSVEHHTNQEDQENNASGSDSDS	SEEGS 442		
Qy	57	SDTSBEGISSDSDLEDEEEDQSIIESEDDSDSEAEQKSNQVLLHGIDSPKAD	116		
Db	443	S-----SNEDSEQNDVDEDDVVESEKRHEPEEGKSSSPGN	-----GHRD-ESN 488		
Qy	117	GOKATEKAEKRIHQPLAFESQTHS	-----FQSOOKOQVLSQQLP-FIFQSSQAK 168		
Db	489	GDKHEDSE-RFQSPSTSSHHETSHSPEKSEAYQSRFSFLNVQSQSPGYEFLESKEI	547		
Qy	169	BESVNKHTSVIQSTGL	-----VSNVKPLSLVNQAKK 199		
Db	548	KQEFPTTSSASSDLEDMEMPDLTRMLERHMRPFIDVSSFVNRIEIVELNQAKR	607		
Qy	200	ETYMKLI-VPSPDVLKAGNKVTSBESSLLTSELRSKREQYQAPP-SOLKQESSKSLKK	257		
Db	608	ASYEKTGRPPFKC-----NNDEVLSIQKIVFHEPRDYVYVYENPCSELEVRIDWRKLS	661		
Qy	258	VIAALSNPKATSSPAHPKQTLNHNHPFLTNALLGNHQPNGVIOQVIAEPLATTKT	317		
Db	662	DTADLDDFRATDSK-----LGRDQAGRTSGRSLDSRTNRL	701		
Qy	318	KMQSKIN-ENIAAAS-----STPF--SSPVNLSTS---GRRTP---GNQTPMP	357		
Db	702	SFDSTHHPAELAQRSLSLIGMPPTSPFPPTSQPLLVNTHLPGTSQBSTSGGITPRSS	761		
Qy	358	SASPILHSQGEKAVNNVNPVKQHHSHHPAKSLVEQPRGTDSIDPSKSEDSNEDEEE	417		
Db	762	QPPPLMSVPSRHNSMSTGRPASIQTLRHQSVMPFPDVSIPPPPPIPTHDEMA--	815		
Qy	418	DEEEDDEDDDESDDSQSDSNSDSTGSEEDDDKQDESDDTEGKTSMKLN	477		
Db	816	-----PRGTPPSRRSS	826		
Qy	478	KTTSSVKSPMSLTHSTPRNLHIAKAPGAPAPALCSQSQAPFLGTSSTLTSSPHSGT	537		
Db	827	ETMVLRSPPF-----GTPIQ-NULLTWPVPPPHLI-----AATSGTHSVSSAHS--	872		
Qy	538	SKRRVTDRELRIPLEYGWORETRIRNFGRLQGEVAYAPCGKKLROYDEVIKYLSRN	597		
Db	873	TPRHSIS-----GTPVHCFPSNSKTSQPTP---KSRP	902		
Qy	598	GIMDISRDNFSAKIRVGDIFYEARDGPOEQMWCLLKEEDVIPRIAMEGRGRPPNPDR	657		
Db	903	EKVQIRHDTIS-----KSGPSN	936		
Qy	658	QRARESRMRKRG-----RPPNVGNAEFLDNADAKLLRLQAEIARQAAQIKLL	708		
Db	937	KKSAPTPVVDAGSDLVIAQIMSNQPNLG-----LRKL--PRIEKKSALQNI	982		
Qy	709	RKLQEQARVAKBAKQQAIAAEERKKQEQIKIMKQOEKIKRIQOIRMEKELRAQOI	768		
Db	983	QNHQPPHSNANSTPSTPSTSTHQAQMFQDKERKK--KEKEKEEREAREEMK-----	1034		
Qy	769	LEAKKKKEEANAALKEAEKRIEKEMRROQAVLLKHQERERRRQHMMLKMAEARKKA	828		
Db	1035	---RKETKEERNKREKEMERAKRLDEQERKRE---KKKERDERK-----KEKEKVRKK	1082		

RESULT 11
A45605
mature-parasite-infected erythrocyte surface antigen MESA - malaria parasite (Plasmodium

Qy	829	BEKERLKOEK-----RDEKRLNKERKLEQORLELEMAKELKKXPNSDMCLAQKPLPE	880		
Db	1083	AEKEKLKKKXHRKGDSDSDSDNBLDVLDRKSTKEMTQEEK-----DHQLA-----	1131		
Qy	881	LPRIPGLVLSGSTSDCLMVVQFLRNFQKVLGFDVNPVLSVLQEGLLNIGDSMGVQ	940		
Db	1132	-----LLLSKGGI-----IENLKKRRSKDKRAHDSFEKQ	1161		
Qy	941	DLLV--RLLSAVCDPGLITGYKAKTALGEHLLNVGNRDNVSEILQIFMEARHCQTALT	998		
Db	1162	QKSQORVLLESSDDEGKDGKNGSNGE-----	1191		
Qy	999	ESLTKAPQAHTPAOKASVLAFLINELACSKSVVSEIDKNIDYMNLRDRDKWVVEGKLK	1058		
Db	1192	ESDSEKADLPPPPAPPS-----LSASADQR	1216		
Qy	1059	LRIIHAKTGKRDPSGGIDLGEEOHPLGTPTPGRKRRKGGDSYDDDDDDSDSDGCD-E	1117		
Db	1217	LKVLKEREKELTTS-----SDDEDHNDAGEIH	1244		
Qy	1118	DDEDEEKQKQKKK--TDICEDE-----DEGQAAASVEE--LEKQIEKLSK	1160		
Db	1245	QORLTEDRENKQKSLTAYSSDEQGERKNVPRMRDDSDAAAKHPGWSAKDDQKQK	1304		
Qy	1161	QSQYRRKLFDASHLSRSMFGPDYRRYRYYWILPRCGGIFVEGMESGEGLEIEIAKEREL	1220		
Db	1305	RKLEHRSSEDESK-----NAKRDPRDIH-----EDVSDEEBTE	1340		
Qy	1221	KKAESVOIKEMPTSGDSLNCSTHCEQKEDLKEKDNVNLFLQKPGSFKSLKLEVA	1280		
Db	1341	DGSR-----RRQSTSTISNVTKERKE-----KSGKTP-----LRIV	1374		
Qy	1281	KMPSESVMPKPNAGANGCTLSYONSGKHSLSGVQSTATQSNVEKADSNLNTGSSGP	1340		
Db	1375	PEPTGTPLSPK-----ILSPKH--LSPKTSSTSKRSISDHNLI-----	1414		
Qy	1341	GKFPYPLNDQLLTKLTKRNOMFSLLPRTCCDTSLTHADMSTASLVTPOSPPSKSPS	1400		
Db	1415	-----SP-----RQRNR-----TTSSTSTATTSSKHEALSIPKPLSPVPTAKS	1453		
Qy	1401	PTAPLPGSSAQNPVGLNPPALSPLOVKG-----GVSMGLOFCQWP	1441		
Db	1454	SVSSIDDPISIRDEFPMNSAADSPMSTTGRPMVLTKAAMKAFNSTPPPKVSYSLIIDC-YM	1512		
Qy	1442	TGVVTSNIPFTLSVPSLGSGLGSEGNGSFLTSNVASSKSESVPQNEKATSQAAPAVE	1501		
Db	1513	LGMMAKNSSGQHDSSGSSSDSSDGS---TSSDSSDDEVP--KQTEPVTISPVVASD	1568		
Qy	1502	VAKP-----VDPPS-----PKPIPE-----EMQFGWRIIDP-	1528		
Db	1569	NGSPENVVETPSIVSQTPREPEPTISEQSSSESEPEAVPECPASVEPQMETSQNVFV	1628		
Qy	1529	-----EDKALLKVLHLRGIREKALQKIQKHLDYITQACLKNK--DVAIELNENENOV	1582		
Db	1629	SEEHED-----SHEHGDSEVAVESQQQPLEHOEKEELEKILDVAA-----BHHEEQV	1677		
Qy	1583	TRDIVENWSVEQAMENDLSVLQVEDLERRRVSASASLQVKGMMCPESAREDLVYFHK	1642		
Db	1678	QGD-----EDSVESI-----PASFDEPDPVTQAEK	1704		
Qy	1643	SFTKLCKE-----HDGEFTGED-----	1659		
Db	1705	SAHTLISDQETDQAVQISIFDEEADPEPQYDFGISTNEKESVKDPHNKIKPTPLNGH	1764		
Qy	1660	-----ESSAHALERKSDNPLDIATVTRLADLERNIERRIEEDIAP	1698		
Db	1765	TDLLFSPSSSAHASEKQSTKSED-----DMEEDSELVVMKEVP	1803		

[illegible]

[illegible]

Qy	829	EKERLKQEK-----RDEKELNKERKLEBQRRLLEMAKELKKNEDMCLADQKPLPE	880
Db	1083	AEEKLKKKXRRKGGDSDESDDSDNDELDDVRKSTKEMTQEBK-----DHQA-----	1131
Qy	881	LPRIPLGVLVSGSTFSDCLMVVQFLRNFVKVLGPDVNI DVPNLVQLGGLNLTGDSMGEVQ	940
Db	1132	-----LLLSKGGI-----IENLKSRRSRDKRAHDSPEKMQ	1161
Qy	941	DLLV-----RLLSAAVCDPLITGYKAKTALGEHLLLVGNVRDINVSEIILQIFMEAHCGQT	998
Db	1162	QKSQQRVLIESDDDEGGDKGNSNGB-----	1191
Qy	999	ESLTKAFAQHTPAQKASVLAFLINELACSKSVYSEIDKNIDYMSNLRDKWYEGKLRK	1058
Db	1192	ESDSEKADLPPPPAPS-----LSASADQR	1216
Qy	1059	LRIIHAKTKGRTSGGIDLGEQHPLGTPTGCRKRRRKGGSDDYDDDDDDSDDDGDD-E	1117
Db	1217	LKVLKEREKGELTTS-----SDDHDNDAGEIHH	1244
Qy	1118	DDEDEDKEDOKKK--TDICEDE-----DEGDOAASVEE--LEKQIEKLSK	1160
Db	1245	QORLTRENRKQKSLTAYSDSEQERKNVPKRRDRDSEDAAKHPGWSAKDDQKQK	1304
Qy	1161	QOSQYRRKLFDAHSLSRVFPGDPRYRRRYWILPRCGGIFVEGMESGEGLEETAKEREKL	1220
Db	1305	RKLEHRRSDESEKK-----NAKRDFRDI PH-----EDVSDEEETE	1340
Qy	1221	KKAESVOIKEMPETSGDSLNCSTHDCQKEDLKEKDNVNLFLQPKGSFSKLS--KLLEV	1279
Db	1341	DGSR-----RRQSTSTISNVTAKEKBE--KSGKTPLRIVPEPTGTLPLSPKILSP	1390
Qy	1280	AKMPPESEVMTKPNAGANGCTLSYQNSGKHSGLSVOSTATOSNVEKADSNLNFNTGSSG	1339
Db	1391	KHLSPKTSSTSKRSSISHENLISPRQRNRTTSST-STATTSKHEALS-----	1439
Qy	1340	PGKFYSPLPNDQLLTKLTEN-----ROWFSLLPRTPCDDTSLTHADMST-----	1384
Db	1440	-----IEKPLSPPTAKSSVSSIDDSIRDEFSM-----NSAADSPMSTGRPM	1484
Qy	1385	-----ASLVTFQSQPPSKS-----PSPTPALGSSAQNPVGLNFPALSP	1423
Db	1485	VLTKAAMKAFNPPPKKETDOAVOSIFDEEEADEFPQYDFGISTNEKEVSGKDPHNIXP	1544
Qy	1424	LQVKGVSVMGLQFCWGPVTGVTNPTFLSVPSLGLSGGLSGEGNSFLTSNVASSKSE	1483
Db	1545	TE-----PLNNGHTDLIFS--PS-----SSHAHEKQSTKSE	1574
Qy	1484	SPVPQNEKATSAO-----PAAVEVAKPDVFPSP-KPIPEEMQFGWRIIDPEDLKALLKVLH	1539
Db	1575	DMDEESELVWEKEVPMQVTAQEVHVSESPMEEEVKLETSPVPKEPIK-----	1627
Qy	1540	LRGIREKALQKIQKHLDYITQACLKNKOVAIIELNENEE--NQVTRDIVENWNSVEQAM	1597
Db	1628	-----MEESPEQTPDLISNSESQDTPGAVNNLHENHDAVQTP	1668
Qy	1598	EMDLVLQOVEDLERRVASASLOVKWCMCEPASEREDLYVEHKSFYCKXEHDEFTG	1657
Db	1669	QLOPASQHQVAQPSRPAPA-----PDSQQNGPVLVSQQS-----QPSPMSS	1710
Qy	1658	EDSSSAHALERKSDNPLDIATVRLADLERNIERRIEDIAPLRVWRRALSEARSAQVA	1717
Db	1711	QOSDMAQNLLSKDINDIA-----AKLKNPE-----ALAQ-----	1743
Qy	1718	LCITQQLQKSIAMEKSTMKVYCOICRKGDNHEELLLCDGCDKGCHTYCHRPKITTIPDGW	1777
Db	1744	-----TRGD-----CSGIFQ--HLLH-----	1758
Qy	1778	FCPACTAKASGQTLKIKLHVKG-----KKTNESKGGKVVTLTGDTEDSASTSSSLKR	1832
Db	1759	-----AONGQNMTPMLQLKAAFPQOQENANOMQAKKQOQITINDKRIKEQSERVK	1812

Db 1157 LDYSTF-----INKLISMQAYNCEFFTDNNGSIT-HVRFARQTPAGHSGKLLNQLPSG 1209
Qy 1464 LSEGNNGSFLTSNVASSKSESFPVQNEKATS 1494
Db 1210 LNDPTATPFTSRPHTSTRASFPFIASSTTQTS 1240
RESULT 15
JC5837
364K Golgi complex-associated protein - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 05-Mar-1998 #sequence_revision 13-Mar-1998 #text_change 09-Jul-2004
C:Accession: JC5837
R:Toki, C.; Fujiwara, T.; Sohda, M.; Hong, H.S.; Misumi, Y.; Ikehara, Y.
Cell Struct. Funct. 22, 565-577, 1997
A:Title: Identification and characterization of rat 364-kDa Golgi-associated protein rec
A:Reference number: JC5837; MUID:98093490; PMID:9431462
A:Accession: JC5837
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-3187 <TOX>
A:Cross-references: UNIPROT:Q63714; DBJ:D25543; NID:G516825; PIDN:BAA05026.1; PID:G5168
C:Comment: This protein plays a role in the formation and maintenance of the characterie
C:Superfamily: giantin
F:49-549,624-1176,1238-1707,1763-3114/Domain: coiled-coil leucine zipper #status predict
F:3165-3187/Domain: membrane anchor #status predicted <MAD>
Query Match 3.2%; Score 326; DB 2; Length 3187;
Best Local Similarity 18.3%; Pred. No. 1.5e-05;
Matches 422; Conservative 362; Mismatches 882; Indels 636; Gaps 88;
Qy 19 EQSNQPLDARDVDIKDKK--PRKAMESSNSDSGTSSTSDTSEGISSDSDDDLEDE 76
Db 747 EQIHSLSIEA---KSKQVKIETLQRELDGVQLQFSEQGTQKSLQSLQTKSEVLEGAE 803
Qy 77 EEPQSEESDDSDSESAHQSNNOVLHGISDPKADQKQATEKAQEKRIHQPLPLA 136
Db 804 RMKDISKEMBELSQALQKLEIAQMDQLL---EKKQVETIQOTIQEKD-QQVTELS 858
Qy 137 PESQTHFSQSOQ-----QPVLSQLPFFIQSSQAKESVNVKHTSVIOSTGLVSNVKP 190
Db 859 FSMTEKVMQLNEEFSLGVEIKLKEQLLSLRAGAKREQV-----EDSGAESPKH 911
Qy 191 LSLVNAQAKETMYKMLIVPSPDLKAGNKVTSSESLTSELRSKREQYKQAFPSQLKKQE 250
Db 912 GPHESSAEEPCKEAL-----QQELEWLKXSEQRKRLQAALTSRKELLQ 957
Qy 251 SSKSLKVVIAALSNPKATSSSPAPHPKOTLE-----NNHPNPPLTNA----- 291
Db 958 KVKLEBELAKVREESTKDSLRESEKRELEBDSKNKDDPEKYGTSEWRELEVSRLTISE 1017
Qy 292 -----LLGNHOPNGVIOQVIEAPALATTKT-----MOSKINENIA----- 328
Db 1018 KEVELEGIRDLKEKAAAEELQALVQMTQDLQNKTKQIDLQEEITENQATIQKFTG 1077
Qy 329 -----AASSTPFPSP-----VNLSSTSGRRTPGNQTPWMPSPAPI 362
Db 1078 TMDAGDGDASVKETSVSPSPRAGGEHWKPELEGIYDLKEBKTQLOKQLQEALISRKA 1137
Qy 363 L-HSQGKEKAVSNVNVFKTHHHPAKSLVQEPFGTDSDIPSSKDSGSDSNEDEEDDEE 421
Db 1138 LKKAQKEKHLKEELKQKDAY-----RHLEQFDG-----QSKENINIRAPLRQ--LQ 1184
Qy 422 EDEBDDDEDDSDQSSESDNSSESDTEGSEEDDDD-----KD 459
Db 1185 AKESTDQOLPGTGOEQEPHGTGSEGLSEGTPEASESDLHAAQPSHPGETATLQATVSAQI 1244
Qy 460 QDE-SDSDTEGEKTSMLNKNTTSVSKSPS-----MSLTGHSTPRNLHIAKAPGSAAPALCSE 515
Db 1245 QDQLKEIYEVEKELELKIISTSTLTKKSEVLLQEQINEQGLEI QNLKAASHEAKAHT 1304
Qy 516 SQSPAFLGTSSSTLTSSPHSGTSKRRRVTDRELRILPLEYGWQRETRIRNFGRLQGEVA 575

Db 1305 EQLKQELSSQLKIADLEHLKTL-----QPELETLQKHVGQKEEVSVLVQLGKEQ 1357
Qy 576 YYAPCGKKLRQYPEVIKYLSRNGIMDISRDNFSAKIRVGDIFYEARDGPOEMQWCLLKE 635
Db 1358 TLTTVTQTEBEEQERLIKAL-----HTQLEMQAK-----EHEERLKQVQVBEICLK 1403
Qy 636 EDVIPRIRAMEGRRRPRPNPDQRARERSMRRRKRRPPNVGNVNAEFLDNADAKLLRLQA 695
Db 1404 QP-----KELEESK-----AKQQLQRLQA 1424
Qy 696 QEIARQAQIKLRLKLOQEQARVAKAEKQQAQIMAAEKKRKOKEQIKIMKQOE-----K 750
Db 1425 ALISRKEA---LKENKSLQEQSSARDAVEHLTKSLAD-----VESQVSVQNEKDALLGK 1477
Qy 751 IKRIQQ-----TRMEKELRAQIIL-----BAKKKKEEAAANAKLLE 786
Db 1478 LALLQEERDKLIVEMDKSLLENOSLGGSCBSLKLALGGLTEDKELMKELSVSRCSIAE 1537
Qy 787 A-EKRIKEEMRRQQAALLKHQE-----RERRQHMMLKAMEARKKAE 829
Db 1538 STEMQEKHLEQKEYEVLLQSYENVSNABERIQHVSVSRQEKQEVVAKURSAESDRER 1597
Qy 830 EKERLKQEKREKRLNKERKL-----EQRLELEMAKELKKPNEDMCLADQKPLPRLP 885
Db 1598 EKQLQDAEQEEMEKEMKWKFAKSKQKILELE-----EENDRLRAEAQPVGANESM 1650
Qy 886 GLVLSGSTFSDCLMVVQFLRNFGLGFDVNDVPLNSVLQEGLLNTGDSMG-EVQDILLV 944
Db 1651 EALLS-----SNASLKEELER-----ITLEYKTLSEKFEALMAEKNTLSEETRLKL 1697
Qy 945 RLLSAAVCDPGLITGYKAKTALGEHLNVGNVDNVSEILQIPMEAHCGOTELTESLTK 1004
Db 1698 QVZAQELKQASLETTEKS-----DEPKDVIIEVTEAVVGSQEQDSUSE- 1741
Qy 1005 AFOAHTPAQKASVLAFILINELACSKSVSEIDKNIDYMSNLRDKWVVEGKRLKRLIHA 1064
Db 1742 --NAKLEDAEATLLA-----NSAKPGVSETFSSHDDINNVLQQLDQLKGRIAELEM-- 1790
Qy 1065 KKTGKRTPSGIGILGEHQHPLGTPTPQKRKRKRGSDSDYDDDDDDDDSDGDEDEDEED 1124
Db 1791 EKQKRELSQTL--NEKNALLTQISAK-----DSELKLEBEVAKINMLNQIQOEL 1841
Qy 1125 KEDQKGGKTDICEDEGDQAAASVEELEKQIEKLSKQSOYRRKLQDA---SHSLRSVMP 1181
Db 1842 SRVTKLKET-----ABEEKDLE--ERLMLQALNGSIGNYQVDTDAQIKNEQLESEM 1895
Qy 1182 GPDYRRRYWILPRCGGIFVEGMESGEGLEIEIAKEREKLLKKAESVQIKKEEMFETSGDSL 1241
Db 1896 N-----LKRC---VSELE--EEKQLVKEKTKV---ESEIRKEYMEKIQQAKG 1936
Qy 1242 CSNTDHC-EQKEDLKEKDNNTLFLQRP-----GSFKLSKLLEVAKMPPESEVWTPK 1292
Db 1937 PGSKIHAKELOQLLEKQEQEVKQKQDCIRYLGRISALEKTVKALEFVHTSESQKDLATK 1996
Qy 1293 PNAGANGCTLSYQNSGKHSIGS--VQSTATQSNVEK--ADSNLFTNGTSSGPGKFYSP- 1347
Db 1997 GNLAQ---AVEHHKKAQELSSFKILLDDTQSAARVLADNLKLLKELQNKESIKSQIK 2053
Qy 1348 -PNDQLIKTL---TEKNR-----QWFSLLRPTPC--DDT-----SLTHADMSTASLV 1388
Db 2054 QKDEDLLRRLLEQAEEKREKKNQKQELDALHREKAHVEDTLAEIQVSLTRKQDKMKEL- 2112
Qy 1389 TPQSPPSKSPSPPTAPLGSSAQNPGVLPNFPALSPQLQKGVSMGLQFCGWPTGVVTSN 1448
Db 2113 -----QQSLDSTLAQAA----- 2125
Qy 1449 IPTLSVPISGLSGLG--LSEGNNGSFLTSNVASSKSESFPVQNEKATSQAAPAAVEVAKPV 1506
Db 2126 --FTKSMSSLODDRDRVIDEAKKWEQFGDAIQTKEEVEVLKEENCTALQKDLQKQMT--- 2180
Qy 1507 DFPSPKPIPEMQFGWRRRIIDPEDLKALLKVL--HLRGIREKALKQKIQKHLDTQACLK 1565

Db 2181 -----IHMEELKITVRLSHDKKEIWESKAQTELOHQKAYDKLOEE 2221
QY 1566 NKDAVAILNE-----NEENQVTRDIVENMSVEEQAMEMDLS-----VLQ 1605
Db 2222 NKEL-MSQLEAGQYHDSKNELTKLESELKSLKQSDTLKNSLEKCREHNNLEGIKQ 2280
QY 1606 QVEDLER-----RVASASL-----QVKGWMCPEPASEREDLYFE-----HK 1642
Db 2281 QEADIONCFNCEQLETDLTASRELTLRLHDEINVEKQIISLLSGKEAIOVAIAELHQ 2340
QY 1643 SFTKLCKEHGDEFTGEDSS-----AHALERKSDNPLDIA 1677
Db 2341 QHSKEIKLENLLSQBEBENLTLEENKRAVEKTNQLETALETIKKESLEKQAK--LDSF 2398
QY 1678 VTRLADLE-----RNIERR-----IEEDIAFGLR----- 1701
Db 2399 VKSMSSQLQDRDRIVSDYRQLEERHLSVILEKDELIQDAAAENNNKKEIR-GLRGHMD 2457
QY 1702 -----VWRRLSEAE-----RSAAQVALCTQIQKSIWESKSMKYVCQICRKG- 1744
Db 2458 LNSNAKLDAELIOYRRDLNEVITIKDSQORQLLEAQQLQONKELRNECVKLEGRLL--KGS 2515
QY 1745 --DNEELLCLCD-----GDKGCHTYCHRPKITTIIPGDWFC----- 1779
Db 2516 EAEKQSLQMSLDALQEEHQGLSKETSP--KEQLTALHEEGALAVYHAQLRVREEVQKL 2573
QY 1780 PACTAKASGOTLKIKKLVKGGKTTNESKGGKVTLTGTDTEDEDSAST-----SSSLKR 1832
Db 2574 TAALSSSQKRTVDLQELVVCQK-----EASKKV-----SEIEDKLKRLKHLHNAGIMR 2624
QY 1833 GNKLOKRMENENTSINLSKQSFVSIVKPKRDDSKOLALCSMLITEMETHED-AMPILL 1891
Db 2625 NETETAERVAELARDLVEMEQLKLTVTYKENDLTAQIAFGKSMSSQLQSDRDHATELS 2684
QY 1892 PVNLKLVPGYKVV--IKKPMDFSTIREKLSQYENLTFALDVLVFDNCETNEDD-- 1947
Db 2695 DLKKYDASLAKELAQKRGDLGRESVLSQAAPLTTSENISSRLEKLNQOLISKQEL 2744
QY 1948 ---SDIGRAGHNMRYFEKKWT 1966
Db 2745 LHLSELESSHNOVQSFTKMT 2766
RESULT 16
D96796
probable heat shock protein, 53413-59028 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004
C:Accession: D96796
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federpiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Frazer, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: D96796
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1871 <STO>
A:Cross-references: UNIPROT:Q9SRD5; GB:AE005173; NID:G6143906; PIDN:AAF04452.1; GSPDB:GN
C:Genetics:
A:Gene: F28016.15
A:Map position: 1

Query Match 3.1%; Score 314; DB 2; Length 1871;
Best Local Similarity 17.9%; Pred. No. 2.5e-05;
Matches 378; Conservative 327; Mismatches 812; Indels 594; Gaps 85;

QY 15 KCNQSQKNQPLDARVD--KIKDKPKPKA-----MESSNSD-----SDSGTSSSTSSSEG 63
Db 59 KIBEEEEEEBERVDVSEAEHKEETEGELKDDYLKSHQIDRIBEEKGLADSNKES 118
QY 64 ISSS--DSDDLE-----EDEEEDQSTEEESDDSD-----SESAQHK----- 100
Db 119 VDSLSRPPDLEGRECHQTRHEBOENKQVQAESDDSDDFGSAFRAFEIEEESDVLDR 178
QY 101 -SNNQVLHGIDPKADQKATERAQEKRIHQPLPLAFESQTHSPSQSQKQPVLSQOQLP 159
Db 179 TSTSGAMEKEMTDVGDGLRVQVQIEEPERHEESKISEMVDGETSGHEKKVV----- 232
QY 160 FIFOSSQAKESVNNKHTSVIOSTGLVSNVKKPLSV-----NQAKETMYMKLIIPS 209
Db 233 ----KMDKKNRDVKEEVDGAMGEGFRPNIDRTQVVGDDEIAETAETKNDEEFES-DKLEADE 287
QY 210 PDVLKAGN-----KNTSESSLLTSELRSKROYKQAPSPQLKQESSKSLKVVIAALSNP 265
Db 288 VDKINEGNTKVRHSEDRNLI--KLOEKEEQH-----SKQKGHSKEENKMKELVEETKP 340
QY 266 KATSSPAHPKQTLNHNHPNFLTNALIGNHPNGVIOSVIQEAPLALTTTKMKOSKINE 325
Db 341 EAEFTT-----IRNDILGPQE-----LEVPEVDLTGKTSDEGKEKQ 376
QY 326 NIA-----AASSTPFSSPVNLSTSGRTPTGNQTPVNPSPASPILHSGQKEKA 371
Db 377 NIVRKEIKNGDATEIIDAEMGEVFA--NIADTGMNSEDFFESDKLESADVDKMKVKKDR 434
QY 372 VSNVNPVKTOHSHHPAKSLVEQFRTGDSIDIPSSKSDSDSNEDEEEDDEDEDEDE 431
Db 435 QEEN-----DKVGAOSEDISLTK-LQEIGEQQFQGGKRDHKOENIKEL 476
QY 432 SDDSQSDSNSDSTEGSEEDDDDKD-----QDESDDSTEG--EK--TSMKLN--- 477
Db 477 REGQASEAKNWKNDILKPVQKRGKHKIQKTQBEINKQPEGNEKIMETGKINEDG 536
QY 478 --KTTSSVKSPPMSLTHGSTPRNLHIAKAPGSAAPALCSESQSPAFLTGSTSTLTSSPHS 535
Db 537 TRKVQEMIRQOELDEPARSEKEN-----RSRELVSKTNDKEEKEIEA 580
QY 536 GTSKERRVTDE----RELRTPLEYQWQRETRINRFGG--RLOGEVAYVAPCGKLLROYFEV 590
Db 581 GTERKEKESDRPKILREQEVADEVA-EDKTKFSYGEVKEEEIEA-----OKEK----- 628
QY 591 IKYLSRNGIMDI SRDNFSFSAKIRVGDYFYEARDPQEMQWCLLKEEDV-----IPR 641
Db 629 -EFGSDDDIARIVDRTQLDSNANKG--EEKMIQE-----LVLEEKVCDGGKGLIAVAE 681
QY 642 IRAMEGRGRPPNPDRQRAEESRMR--KGRPPNVGNABFLDNADAKLRKLQAO 696
Db 682 TKAENNNKSRVQETEEQKLDKEDTCGKHFOKLIEGEISDHGEVDVEKGGKRTAEKRIK 741
QY 697 BIARQAAQIKL--LRKLOKQEOARVAEAKKQQAAMAAEKKKOK-----EQIKMKQO 748
Db 742 DRAREABEIEKDLGVSGRYIKGTTIKELVENRGIYRNEHEEKKDDANRPEKITGTIKQ 801
QY 749 EKIKRIQOIRMEKEL---RAQOILEAKKKKEEANAALLAEKRIKEKEMRROQAVLKK 805
Db 802 ELVSLNQLQKQENVEDGDKTQELVEEKIKOCEESEGE-----ESKIKTDDVVVRQVQGIKE 857
QY 806 HQERERERQH-----MMLMKAMEARKKAAEKERLQKQRDE--KRLNKERKLEQRRL--ELE 858
Db 858 EELYKPKREHGTKITELVEETTGDEYEQEKEETAESDIEAECGSLRVKVDGIEEHELHPK 917
QY 859 MAKELKFPNEDMCLADQKPLPELPRIPLGLVLSGTSFSDCLMVVQFLNFGKVLGDFDVID 918
Db 918 IHKE--RDNNRVGTAKFPGSQGKEKEKIVESMTITE-----NDNSID 959
QY 919 VPNLVLQEGLLNIGDSMGEVQDILLVRLLSAAVCDPGLITGYKAKTALGEHLNVGVNRD 978
Db 960 VQETKKEPGRLESHDKRYKIQLL----- 984
QY 979 NVSEILOQIFMEAHCGQOTELTESLTKFAQHTPAQKASVLAFINELACSKSVVSEIDKN 1038


```
Qy 523 GTSSSTL--TSSPHSGTSKRRVTDRELRIPLEY-----GQRETRIRNFCGRLOQ-EVA 575
Db 1173 NTQLOSLQAQSIKSLDQPKQDLQEEGVKLLLEWELLKGLHTDSQISIEKLOJLENLEVT 1232
Qy 576 YYAPCGKKLQYPEVIKYL--RNGIM----DISRDNFSFSAKI--RVGDFYEARDGPOE 627
Db 1233 -----EKLQTLQEMKNITIERNELQTNFEDLKAHDSLQDLSENIEQSIETQDELRA 1286
Qy 628 MOWCLLKEDVPIRIR--AMEGRIG-RPPNDRQARF-----ESMRRRKGRPP 674
Db 1287 AQBELREQQLVDFRQQLLDCSVGISPNNDVAVANQEKVSLGVNSLQSEMLR----- 1340
Qy 675 NVGNAEFLDNADAKL---LRKLAQAEIARAQAQKLLRKLQKQQAQVAK-----EAKKQQ 727
Db 1341 --GERDELQTSCKALVSELELRHVKVSEGENEITKKUNGLEKEITLGKSESESVLKSM 1398
Qy 728 AIMAAEKKRQKEQIKIMQOEKIKRIQOI--RMEKELRAQOIILEAKKKKKE----- 778
Db 1399 LENLKEDNNKLEQAEEYSSKENQFSLSEVFGSQKLVDLEVLKAQLKAAERLEIKDR 1458
Qy 779 -----AANAKLLEA-----EKRIKEMR-----RQAVL--LKHQ 807
Db 1459 DYFELVQTANTLVGKLETPLOADHEEDSDRSEEMEIKVLGKELRNOYLLERLQEE 1518
Qy 808 ERERRRQHMLMKAMEARKKAEERLQKEK-RDEKRLNKE-----KLEQR 853
Db 1519 KLELSNKLEILQKEMETSVLKDQKLESLSENILKENIDTTLKHSDDTAQLOKT 1578
Qy 854 RLELEMAKELKKPNEDMCLADQKPLPELPIPGVLVSGSTPDCILMVVQFLRNFGKVLGF 913
Db 1579 QOELQAKNLAIASDNCPIQEK-----ETSADC----- 1608
Qy 914 DVNIDVNLVQLQGLNLIGSMGEVQDLLVRLLSAAVCDPLGTGYKAKTALGEHLNV 973
Db 1609 -----VHPLBEKILLLTTELHQ-----KTNEQEKLLH-- 1635
Qy 974 GVNDRNVSEILQIFMEAHC-----GQTELTESLTKTAKFOAHTPAQKASVLAFINELA 1026
Db 1636 -----EKNELEQAQVELKCEVHELMKSMIESKSLSESLQHEKHDTQOALLALKQOMQVVT 1690
Qy 1027 CSKS-----VVSEID---KNIDYMSNLRD-----KW 1050
Db 1691 QEKKELQOTHEHLTAEDVHLKENIELGNFKNEAQAQKTTKEQCLLNKENLEQSQHRLQC 1750
Qy 1051 VVEGKLRLRIH-AKTGKRDTSGGIDLGEQHPLGTPPPGRKRRKGGSDVDYDDDD 1109
Db 1751 EIEELMKSLKDKEALETLKSEQKVINLQEMEMVMEELKNSQRTVIAERDQLQDD 1810
Qy 1110 --DSDQDQDEDED--EEDKEDQKGKKTIDICEDEGEDQAASVEE---LEKQI--- 1155
Db 1811 LRESVMSIEFTQDLRKAQEAALQOQKQKVQLT-----SQISVLQEKISILENQMLNV 1864
Qy 1156 ----EKLSKQO--SQYRRKLPD----ASHSLRSVMFG-----PDRYRRRYWILPRCG 1197
Db 1865 ATVKETLSERDDLQSKQHLFSEIETLSLSLKEKEFALEQAQKADAAKTIIDITEKIS 1924
Qy 1198 GIFVEGMEGGGLBEIAKAREKLKKAESVQIKEMFETSGDSLNCNSNDHCEQKEDLKEK 1257
Db 1925 NIEBQLLQQAATNLKETLYERESL-----IQCKEQL-----ALNTEHL--RETLLSK 1968
Qy 1258 DNT--NLFLQKPGFSKLSKLLVAKMPPESE-----VMTPKPNAGANGCTLSVQNSGKH 1310
Db 1969 DLALGKMEQERDEANKVIALTE--KMSLEEQINENVTTLKEGEGSKE-TFYIQRPSKQ 2025
Qy 1311 SLGVSQSTATQS-----NVEKAD-----SNNLFTNGSSGPGKGFYSPFLPNDQLLKLITE 1358
Db 2026 QSSQMEELRESLTKDQLQLEAEAKEISEATNEIKNLTAK-----ISSLEE 2071
Qy 1359 KNRQWFSLLPTPCDDTSLTHADNSTASLVTPQSQPPSKSPSPPTAPLGGSAQNPVGLNP 1418
Db 2072 EILQNASITLNAEVSERENLRH---SKOQLVSELEQLSITLKSRDHAFQAQSKREXDEAVNK 2128
```

```
Qy 1419 FALSPLQVKGVSMMGLQFCGWPTGVVTSNI-PFTLSVPSLGSLGSLSEGNNGNSFLTSNV 1477
Db 2129 IASLAEBIK-----ILTKEMDEFDRDSKESL-----QEQSSHLSSEEL 2164
Qy 1478 ASSKESBPQNEKATSQAQAAVEVAKVPDPFPPSPKPIPEEMQFGWRIIDPEDLKALLKV 1537
Db 2165 CTYKTELQMLKQKQKEDINNKLAEKV-----KEVDELLQ----- 2197
Qy 1538 LHLRGIREKALQKQIQ-----KHLDYITQACLKNKDVAIELNENENOVTRDIVENWS 1591
Db 2198 -HLSLKEQDLQIQMELRNKRLNYELCEKNDIMEKISVLRLMQNEPQEQEEDDVAERMD 2256
Qy 1592 VEEQAMEMDLSVLQOVDELERRVASASLOVKGWMCPEPASEREDLVYFEHKSFTKLCKEH 1651
Db 2257 ILE-----SRQEIQELMEKISAV-----YSEQHTLLSSLSSELOKETEAH 2297
Qy 1652 DGEFTGDESSAHALER-----KSDNPLDIAVTRLADLERNI-----ERRIEDIA 1697
Db 2298 KHCMLNIKESLSSSTLSRSFGSLQTEHVKLNTQTLTLNKFVVYRTAAVKEDHSLIKDYE 2357
Qy 1698 PGLVWRRLSEARSAAQVALCTQOLOKSIAWEKSINK--VYCOICRKGDNELLLLLCDG 1755
Db 2358 KDLAAEQKHDELRLQLQ---CLEQHGK--WSDSASEELKFCEI--EFLNELLF----- 2405
Qy 1756 CDKGCHTYCHRPKITTIPTDGDWFCPACIAKASGOTL-----KIKKLHVKG 1800
Db 2406 -----KXANIIOVSQDDFSEVQVFLNQVSTLQEELEHKKGFQMWLEEFQDLHVDA 2456
Qy 1801 KKTNESKKGKVTLTGDTDEDEDSASTSSSLKRGKNDLOKRMBSNTSINSKQBSFTSVK 1860
Db 2457 KKLSEGMQ-----QENRRITASTIQLTLKRLKAVVQSKIOREITVYLNQ---FEAKL 2504
Qy 1861 KPXRDDSKDLALCSMLITEMETHEDAWPFLPLVNLKLVPGYKKVKKPMDPFSTIREKLSS 1920
Db 2505 QEKKEQNKL-----MRRMEHGHGPSASVMEENARLL-----GILKTVQDES---KKLOS 2551
Qy 1921 GOYPNLETFALDVLRLVFDN 1939
Db 2552 ----RIKMLENELNLVKDD 2566

RESULT 19
A45973
trichohyalin - human
C;Species: Homo sapiens (man)
C;Date: 03-May-1994 #sequence_revision 01-Mar-1996 #text_change 09-Jul-2004
C;Accession: A45973
R;Lee, S.C.; Kim, I.G.; Marekov, L.N.; O'Keefe, E.J.; Parry, D.A.D.; Steinert, P.M.
J. Biol. Chem. 268, 12164-12176, 1993
A;Title: The structure of human trichohyalin. Potential multiple roles as a functional E
ed (cross-linking) protein.
A;Reference number: A45973; MUID:93280194; PMID:7685034
A;Accession: A45973
A;Molecule type: DNA
A;Residues: 1-1898 <LEE>
A;Cross-references: UNIPROT:Q07283; GB:L09190; NID:g292835; PIDN:AAA65582.1; PID:g292836
A;Note: authors translated the codon AGG for residue 1714 as Pro
C;Comment: Trichohyalin is a protein of the medulla of the hair and of the inner root sh
Covalent modifications to this protein include conversion of arginine to citrulline and
C;Genetics:
A;Gene: GDB:THH
A;Cross-references: GDB:136223; OMIM:190370
A;Map position: 1q21-1q21
C;Superfamily: trichohyalin; calmodulin repeat homology
C;Keywords: calcium binding; citrulline; EF hand; hair; tandem repeat
P;49-81/Domain: calmodulin repeat homology <EF2>

Query Match 3.1%; Score 309.5; DB 1; Length 1898;
Best Local Similarity 18.0%; Pred. No. 3.9e-05;
Matches 185; Conservative 201; Mismatches 391; Indels 251; Gaps 32;
Qy 394 QFRGTSDSDIPSSKDSKSDS--NEDEEEDDEDEDEDEDDSDSDSQESDSNSESDDTGESE 451
```


Db 1554 EK-TIDSEKDLLDDSD-----DDDIIEILBECIISAMPTKSSRKAKKPAQTAASKLPPPV 1605
QY 489 SLTGHSTP-----RNLIHAKAPG-SAPAAALCSSESQSAFLGTSS-----TLTS 531
Db 1606 ARKPSQLPVYKLLPSQNRLOPKQHVFTPGDDMRVYCVBG-TPIINFSTATSLDLTIES 1664
QY 532 SPH-----SGTSKRR-----RVTDERE-----LRIPLEYGWQRETRIRN 565
Db 1665 PPNELAAGEVGRGNAQSGEFKRTIPTGEHSTDEAOGGKTSSVTIP-----ELDDNKAE 1719
QY 566 FGRLQGEVAYVAPCGKKLROYPEVIKYLSRNGIMD-ISRDNFSAKIRVGDIFYEARDG 624
Db 1720 EGDILAEICINSAMPKSKHPF-RVKK-----IMDQVOQASASSA----- 1759
QY 625 PQEMQWCLLKEDVIPRIAMEGRGRPPNDR---QBARSESRRKRGRPPNVGNAE- 680
Db 1760 -----PNKNQLDGGKKKPTSPVKPIQNTYRTRVRKNADSKNL-NAER 1803
QY 681 -FLDNADAKLLR-----KLOAOE----- 697
Db 1804 VFSNKKUSKKQNLKNSKDFNDKLPNNEDRVGRSFAFDSPHHYTPIEGTPYCFSRNDSLS 1863
QY 698 -----IARQAQIKLRLKQEQARVA-----KEAKQQAIAAAEERKQ 738
Db 1864 SLDFDDDDVDLSREKAELRKAKE-NKESEAKVTSHTELTNSQOSANKTQAIKAPINRGQ 1922
QY 739 KEQIKIMKQO-----EKIKRIOQIRME-----KEIRAQOILAEAKKK 775
Db 1923 PKPI-LOKQSTFPOSSKDIPRGAATDEKLQNFIAIENTPVCFSHNSLSLSLSDIDQENNN 1981
QY 776 KEEAANKLLEAKRIKEKEMRRQOAVLLKHOERERRRHQMLMKAMEARKKAEKER-- 833
Db 1982 KE---NPIKETEPDPSGEGSKQA-----SGYAPKSFVETDTPV 2019
QY 834 --LKQERDEKRLNKRKLEQRLEMAKELKKPNEDMCLADQKPLPELPRIPLGLVLSG 891
Db 2020 CFSRNSLSLSIDSEDDLQECISSAMPKK-KKPSR-LKGDNEKHSP----- 2065
QY 892 STFDCLMVQFLBNFGKVLGFDVNDVPNL-----SVLQGLNLIG 933
Db 2066 -----RNMGGILGEDITLKDQIRPDSEHGLSPDSENFDMKAIQEGANSIV 2112
QY 934 DSMGEVDLLVRLLSAAVC-----DPGLITGVYKAKTALCEHLNVLGNVRNDVSEILQI 986
Db 2113 SSLHQA-----AAACLRSQASSDSLSLSKSGISLGSFP----- 2148
QY 987 FMEAHCGQTELSLTKTAFOAH-----TPAQKASVLAFILNELACSKSVVSEIDKNID 1040
Db 2149 -----HLTPDOEEKPFTSNKGPRILKPGEKSTLE-----TKKIESE----- 2184
QY 1041 YMSNLRDKWVVEGKLKRLRIHAKYTKGRTSGDILGEBQHPGLGTPPGKRRKRGD 1100
Db 2185 -----SKGIKGGK---KYVKSLLITGKVRNSEIS-GQMKQPLQANMPSISGR--- 2228
QY 1101 SDYDDDDDDDDGDEDEDEKQKKG-KTDICEDDEGDQA-----ASVELEK 1153
Db 2229 -----TMHIHFGVNRSSSTSPVSKGKGPLKTPAKSPSEGTATTSPRGAKPSVKS 2280
QY 1154 QIEKLSQQSQYRRKLFDAHSLRSVMFGPDYRYYRYWILPRCGGI FVEGMESGEGLEEI 1213
Db 2281 ELSFVARQTSQ-----SRDS-----TPSRPAQQLSRP-----IQSPGRNSIS 2330
QY 1214 AKEREKUKAESVOIKEMFTSGDLSLNCSTNDHCEQKEDLKEKDNNTLFLQKPGSPS-- 1271
Db 2304 -----SRDS-----TPSRPAQQLSRP-----IQSPGRNSIS 2330
QY 1272 ---KLAKLLEAVKMPPESEVMTPKPNAGANGCTLSYQNSGKHSILGVSQSTATQSNVEKA 1327
Db 2331 PGRNGISPPNKLSQLPRTSSPTASTKSSGSG-KWSTYSPGRQM--SQQNLTKGTGLSKN 2387
QY 1328 DSN-----NLFNVTGSGGPKFYSPPLPNDQLLTKLTEKNRQWFSLPRTPCDDTS 1376
Db 2388 ASSIPRSESASKGLNMNGNGANKKVE-----LSRMSSTKSS 2425

QY 1377 LTHADMSTASIVTQSQPPSKSPSPT-----PA-PLGSSAQNPPVGL 1416
Db 2426 GSSDRSERPVLVRQSTFIKEAFSPTLRRKLRESASFESLSPSRPASPTRSQATPV-- 2483
QY 1417 NPFALSLQVKGGVSM-MGLQFCGWPTGVVTSNIPFTLSVPSLGSGLSGNGNSFLTS 1475
Db 2484 ----LSFSLPMSLSLTHSSVQAGW-----RKLPPNLS-PT-----IEYNDGRPAKRH 2526
QY 1476 NVASSKSESP--VPONEKAT-----SAQPAAV 1500
Db 2527 DIARSHSESPRLPINRSGTWKREHSHSSSLPRVSTWRRRTGSSSSILSASSESEKAKS 2586
QY 1501 EVAKPVDFPSKPIPEEMQF---GWWRIIDPEDLKALLKVLH-----LGIKREKALQKQ 1551
Db 2587 EDEKHVNSISGTQSKQSNQVSAKGTWRKIKENEFSPNTSQTVSSGATNGAESTLIYQ 2646
QY 1552 IQKHLDYITQACLKNKDVAI-IE-----LNENEE 1579
Db 2647 -----MAPAVSKTEDVWVRIEDCPINNPRSGRSPGTGNTPPVIDSVSEKANPNIKDSK 2699
QY 1580 NOVTRDIVENWVBEQ--AMEMDLSVLQOVEDLERRVASASLQVK-GWMCPEPASEREDL 1636
Db 2700 NOAKQN-VNGSVPMRTVGLNRLNSFIQVDAPDK---GTEIKPGQNNPVPVSETNES 2754
QY 1637 VYFEHKSFT 1645
Db 2755 SIVERTPFS 2763

RESULT 23
S33124

tpr protein - human
N;Alternate names: kinase-related transforming protein (tpr-met); protein with promoter
C;Species: Homo sapiens (man)
C;Date: 22-Nov-1993 #sequence revision 26-May-1995 #text_change 09-Jul-2004
C;Accession: S33124; S23740; S00928; G01185
R;Mitchell, P.J.; Cooper, C.S.
Oncogene 7, 2329-2333, 1992
A;Title: The human tpr gene encodes a protein of 2094 amino acids that has extensive coi
A;Reference number: S33124; MUID:93064711; PMID:1437155
A;Accession: S33124
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: mRNA
A;Residues: 1-2094 <MIT>
A;Cross-references: UNIPROT:Q15624; UNIPROT:Q9UE33; EMBL:X66397; NID:G633225
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1991
R;Mitchell, P.J.; Cooper, C.S.
Oncogene 7, 383-388, 1992
A;Title: Nucleotide sequence analysis of human tpr cDNA clones.
A;Reference number: S23740; MUID:92195670; PMID:1549355
A;Accession: S23740
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-725, 'L' <MI2>
A;Cross-references: EMBL:X63105; NID:G37257; PIDN:CAA44819.1; PID:G37258
R;King, H.W.S.; Tempest, P.R.; Merrifield, K.R.; Rance, A.J.
Oncogene 2, 617-619, 1988
A;Title: Tpr homologues activate met and raf.
A;Reference number: S00928; MUID:88262257; PMID:3387099
A;Accession: S00928
A;Molecule type: mRNA
A;Residues: 1-31, R', 33-142 <KIN>
A;Cross-references: EMBL:Y00672; NID:G37255; PIDN:CAA68681.1; PID:G37256
R;Greco, A.
submitted to the EMBL Data Library, December 1995
A;Reference number: H00592
A;Accession: G01185
A;Status: translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 144-228 <GRE>
A;Cross-references: EMBL:X94208; NID:G1296797; PIDN:CAA63904.1; PID:G1296798
C;Genetics:

A:Gene: GDB:TPR
A:Cross-References: GDB:128821; OMIM:189940
A:Map position: 1q25-1q25
A:Introns: 177/3

Query Match 3.0%; Score 301; DB 2; Length 2094;
Best Local Similarity 18.0%; Pred No. 9.8e-05;
Matches 353; Conservative 329; Mismatches 743; Indels 536; Gaps 78;

```
QY 15 KNCQSQSK-----NQPLDARDV-----KIKKKPKKAMSSSSND----- 50
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 223 KCNLENKKEEVSRLBEQNGMLKTSNEHLQKHVEDLLTKLKEAKEQOASMEKFFHNLNAH 282
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 51 -----SDSTSSDTSSEGI-----SSSDSDLEDEBEDEQ----- 81
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 283 IKLSNLYKSAADDSAKSNELTRAVEHLKLLKEAGEANKATQDHLLEVEQSKQDMKEM 342
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 82 -----SIESEDDSD-----SESAQHSNNQVLLHGHISDPKADQKAT--- 121
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 343 LEKIGRLEKELENANDLLSATYRKGAIISEELAAMSPATAAAKIVKP---GMKLTLY 399
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 122 -----EKAQEKRIHOPL-----PLAFESQTHSFQSQKQPOVLSQL 158
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 400 NAYVETQDQLLEKLENKRINKYLDIEIVKEEAKAPI-LKQREBEYERAKAVASLSVKL 458
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 159 PFIFOSSQAKER---SVNKHTSVIQ---STGLSVNVKPLSLVNOAKETMYKMLIVSPDVL 213
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 459 EOAMKEIQRLOEDTDKANKQSSVLSDNRMRMEIQVKDLS-----QQIRVLLMELE 508
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 214 KAGNKNTSESSLLTSELRSKREQYKQAPFQPKQKQESSKSLKKVIAAL-----SN 264
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 509 EARGNHVIRDEVSADSSSSSEVTSQHLVSVYNTIEEQOQNRLLVALRELGETRERE 568
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 265 PRATSSSPAHKQTLNHNPNPFLTNALGNHQPNGVTSQVIEAPLALTTKTKMQSKIN 324
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 569 QETTSKITELQKLE-----SALTELEQLKRSQHQHQLVDSIVRQDMVRILLS 619
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 325 ENIAAASSTPSS---PVNLSTSGRTPGNQTVPWSPASPIIHSQKQEKANVNVNPKVTK 382
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 620 QTTGVAIPLHASSLDVSLASTPKRPSQTVSTPAPVPIESTEAIEAKA----- 670
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 383 HHSHPAKSLVEQFRCTDSDISSKSDSESDNEDEDEDEDEDEDESDSSESNSN 442
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 671 ----ALKQLEIF-----ENYKKEAKENKIQNEQKLEQEQVTDL 707
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 443 SESDTGSEEDDDKDDODESDSDTEG---EKTSM-----KLNKTTSSVKSPSLTCHS 494
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 708 RSQNTKISTQDLPASKRYEMQLQDNVEGYRREITSLHERNQKLTATTQKQEIINTWQDL 767
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 495 TPRNLHIAKPGSAP-----AALCSRSQSPAFIQTSSSTLTSSPH 534
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 768 RGANEKLAVAEVAENLKKKEMLKLSVRLSQQRESLLAEQGNLLLTNLQTI----- 822
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 535 SGTSKRRVTVDERELRIPLVQWQRETRIRNFGRLQGEVAYAPCGKKLQYPEVIKYL 594
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 823 QGILSERSETQRILSSQIE-----KLEHIEHLK---KKLENEVEQRHTL 865
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 595 SRN---GMDISRD-----NFSFSAKIRVDFYEARDGPQEMOWCLLKEEDVPIRAME 646
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 866 TRNLVDQLDTRQLDTRTNLHNTKELL-----KNAQKEIATLKQHSNNEVQVASQS 919
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 647 GRR---GRPPNPD-----RQARBEESRMRR-KGRPPNVGNAE-FLDNADAKLARK 692
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 920 SORTKGQPSNKEDVDLVSLQRQTEQVNDLKERLKTSTSNVEQYQAWVTSLEESLNKE 979
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 693 LQAQEIARQAQIKLLR---KLQKQEARVAEAKQQAIAAEBKRK-----QKQIKIMK 746
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 990 KQVTEEVAKNTEVLKESABFQTLEKKLMEVEKEQELQ---DDKRAIESMEQQLSELK 1037
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 747 Q-----QEKIKRI-----QQIRMEKELRAQQIILEAKKKKEB-----AANAKLE 786
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1038 KTLSSVQNEVQEARQASTALSNEQQARRDCQEQAKIAVEAQNKYTERELMLHAADVEALQ 1097
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

RESULT 24

S28261

Centromere protein E - human

N:Alternate names: centromere 312K protein; kinesin-related protein CENP-E

```
QY 787 AEKRIKEMRRQQAALLKHQERRRRROHMMMKAMAEARKKAEBEKERLQKQKDRKLNK 846
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1098 A-----AKEQVSKWASVRQHLDETQKAESQL-----LECKASWEERERMLKDESKVCVCR 1149
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 847 ERKLEQRELEMAKELKKPNEDMCLADQKPLPELPRIPGLVLSGSTFSDCLMVVOFLRN 906
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1150 EDLEKQNLHLDQIEKL-----SDKVVASVKGVGQPLNVSSEKSGOEILIEURFIRR 1205
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 907 FGKVLGFDVNDVNPNSVLQ-----ODLLVRLLSAAVCDPGLITGYKATALGE----- 968
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1206 EKEITA--ETRFEVAVESLRVQRVVELLERELQLELSDLSNAERKQVTTAKTMAQHEELM 1263
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 931 NIGDSMGV-----ODLLVRLLSAAVCDPGLITGYKATALGE----- 968
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1264 KKTETVMVMTNKMRLBEKERLEQLQMQAKVRKLELDILPLQEAANAELSEKSGMLQA 1323
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 969 --HLLNVGVNRDNVSEILQIFMEAHCGQTELTESIKTAKFOAHTPAQKASVLAFILNELA 1026
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1324 EKKLEEDVKKWAKARNQHLVSQLQDPDTEEFYRKLSEKEVHTKRIQQTTEEIGRLKAEITA 1383
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1027 CSKSVSEIDKNIDYMSNLRDRKWVEGKRLKRLIIHAKKTGKRDTSGG-IDLGEEQHPL 1085
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1384 RSN---ASLTNNQNLIOQLKED-----LNKVRT--EKETIQOLDAKIIDI---QEKV 1428
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1086 GTPTPGRK--RRRKGSDSDYDD-----DDDDSDDDQDEDEDEDEEKE--DQ 1128
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1429 KTIITQVKIGIRRYK--TQYEELKAQDKVMETSAQSSGDHQEHQVSVQEMQELKETLNQ 1485
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1129 KGKKTIDICEDE-----GQAAAVEELEKQIEKLSKQSOQYRKLKFDAS--HSLRSV 1179
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1486 AETKSLSQVENLQKTLSEKETEARNLQEQTVQLOSELRLRLQDLQDRTTQEBQLRQ 1545
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1180 MFGPRDRYRRYWIILPRCGGIFVEGMESEGELEEETAKEREKLKKAESV--QIKKEE-- 1232
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1546 ITEKEEKTRKAIKAAKSKIAHLGVK-----DQLTKENEELKQNGALDQKDELVDVIT 1600
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1233 ----FETSGDSLNCNDHCEQEDLKKEKNTNLFQKPGSFKSLKLELVAKNPPSE 1287
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1601 ALKSQYEGRISRLRELREH--QERHLEQRDEP---QEPSN-----KVPEQQR 1643
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1288 VMTPK--PNAGANGCTLSYQNSGKHSGLSVQSTATQSNVEKADSNLNTGSSGPGKFYS 1345
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1644 QITLKTTPASGERGIA-----STSDPPTANIKPTPVVSTPSKV-----TAAAMAGNKST 1692
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1346 PLPNDQLLKLTEKNRQWFLPRTPCDDTSLTHADMSTASLVTPOSPPSKSPSPPTAP 1405
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1693 PRASIRPMT-----PATVNTPTTPTATVMTPTTQVESQEAQSEGPEVHPVP 1740
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1406 LGSSAQNPNVGLNFPALSPLOVKGVS--MMGLQFCGMPTGVVTSNIPFTLSVPSLGSLG 1463
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1741 FGSTSGSVRSTSP-----NVQPSISQPLTVQOQTQATAFVQ---PTQQSHPOIEP--- 1788
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1464 LSEGNHNSFLTSNVASSKSSSPVPONEKAT-----SAQPAAVEVAKPVPDPSPKPIEE 1517
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1789 ----ANQELSSNIVEVQSSPVERPSTSTAVFTVATSPSS-----SLPKRTREE 1834
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1518 MQFGWRIID-----PEDLKALLKVLHLRGIREKALQKQIKHLDYITQACLNK 1566
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1835 EEDSTIEASDOVSDDTVEMPLPKLKSVTPV-----GTEEEVMAEE-----STD 1878
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1567 KQVAIELNEENENQVTRDI-----VENWMSVEQAMMDLSVLQVQVEDLERRVASASL 1619
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1879 GEVETQVYNQDSQDSIGEGVTQGDYTPMEDSEETSQSILDLGLPLQ--SDQQTTSQDG 1936
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1620 QVKGMMCPPEPASREDLVYFEHKSFTKLCKEHDGCFTEGE 1660
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1937 QGKG-----DDVIVIDSD-----EEDEEDDDDDDE 1962
||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

C:Species: Homo sapiens (man)
C:Date: 03-Mar-1994 #sequence_revision 03-Mar-1994 #text_change 09-Jul-2004
C:Accession: S28261
R:Yen, T.J.; Li, G.; Schaar, B.T.; Szilak, I.; Cleveland, D.W.
Nature 359, 536-539, 1992
A:Title: CNP-E is a putative kinetochore motor that accumulates just before mitosis.
A:Reference number: S28261; MUID:93024922; PMID:1406971
A:Accession: S28261
A:Molecule type: mRNA
A:Residues: 1-2663 <YEN>
A:Cross-references: UNIPROT:Q02224; EMBL:Z15005; NID:g29864; PIDN:CAA78727.1; PID:g29865
C:Genetics:
A:Gene: GDB:CNPE
A:Cross-references: GDB:361164; OMIM:117143
A:Map position: 4q24-q25
A:Superfamily: centromere protein E; kinesin motor domain homology
C:Keywords: ATP; coiled coil; microtubule binding; mitosis; nucleotide binding; P-loop
F:7-335/Domain: kinesin motor domain homology <XMO7>
F:86-93/Region: nucleotide-binding motif A [P-loop]
F:486-2183/Domain: coiled coil #status predicted <COI>
F:92/Binding site: ATP (Lys) #status predicted

Query Match 3.0%; Score 300.5; DB 1; Length 2663;
Best Local Similarity 18.9%; Pred. No. 0.00014;
Matches 380; Conservative 308; Mismatches 711; Indels 611; Gaps 89;

QY 15 KCNQEQSQKNOPLDARVDKIKDKPKKAMSSNSDSGCT-----SSDTSSEGI 64
DB 871 KTQELQKTRVQERLNMQLK-----EQLENRDSPLQTVREKTLTEKLOQTLEEV 924
QY 65 S--SSSDDDLEDEE-----BEDQ-----SIESEDDSDSESEAQHKSNNQVLL 107
DB 925 KTLTQEKDCLKQLESQIERDQLKSDIHDTVMNNDITQELRNALLESKQHOETINTLK 984
QY 108 HGISDP-----KADQKATEKAQEKRIHQIPLAPESQ-THSFOSQKQPOVLSQLP 159
DB 985 SKISEEVSRLNMBEENTGETKFOQRWQIDKKQDLEAKNTOTLTADVXNDNIEEQRK 1044
QY 160 FIFQSSQAKESVYKHTSVIQ-----STGLVSNVKKPLSLVNO-----AKETVMKL 205
DB 1045 -IFSLOQEKNELQOMLESVAEKEQLKTDUKENIE-MTIENQEBELRLGDELKKQEI 1102
QY 206 IVFSPDVLKAGN-KNTSEESLLTSELRSKREQYKQAFPSQLKKQBSKSLKXVIAALS 264
DB 1103 QEKNHAIKKEGELSRCTDLAEVVEKLEKSKQLQEKQQLLVQEMSEWQKINEIEN 1162
QY 265 PKATSSPAHPKQTLNNHNPNFLTNALLGNHOPNGVIOQVIOEAPIALTTKTMOSKIN 324
DB 1163 -----LKNELKNKELTL-----EHME-----TERLELAQKLN 1189
QY 325 ENIAAAASSTPFSPV-----NLSTSGRRTPGNTQPMPSASPIHLSQGEKAVSNVNP 378
DB 1190 ENTEEVKSTYKERKVLKELQKSPETERDHURGYRIEATG-----LQTKEE----- 1236
QY 379 VKTQH-HSHPAKSLIVEQFRGTDSD----IPSKDSDESDNEDEDEDEDEDD----- 427
DB 1237 LKTAHILKHEQETIDELRSVSEKTAQIINTQDLEKSHTKLQEEIPVLHEEQELLPNVK 1296
QY 428 EDESDDSQSESQSNSESDT-----EGSEEDDDDKQDESQSDTGEKTSMK 475
DB 1297 KVSETQTMNELELLTTSQSTKDDSTTLARIEMERLRNLEKRFQSQBEIKSLTK-ERDNLK 1355
QY 476 LNKTTTSVKFSPMSLTGHSFPRNLHIKAKGSAAPALCSQSQPAFLGTSSTSLTSSPHS 535
DB 1356 TIKEALEVGHQL-----KEHRET-----LAKIQESQSKQ---EOSTLNMKEKDN 1398
QY 536 GTSKRRRVTDRELRIPLYEGWRQ-ETRIKRNFGGRLO---GEVAYVAPCGKKLRQYPEVI 591
DB 1399 TT-----KIVSEMEQFKPKDSALLRIETIEMGLSKRLQESHDEMKSVAKEDDLQRLEVL 1454
QY 592 KYLSRNGIMDISRDNFSFAKIRVGDYFYEARDGPQEQMWCLLKKEEDVIPRIRAMEGRRGR 651
DB 1455 QSESF-----DQKENTKEIVAKHLETEEBELKVAHCCLEQKEQETINELRNLS---- 1501

QY 1532 KALKVHLHIGIRE-KALQKQIQKHLDYITQACLKNKDVAILIENEN-BENQVTRD----- 1585
 Db 2423 KMKLVKIDLEKMKNAKEFEKES-----ATKATVYQKEVIRLLRENLRSSQAQDTSVI 2477
 QY 1586 -----IVENMSV-----EEQAMEMDLS-----VLOQVEDLER 1612
 Db 2478 SEHTDPQPSNKPLTCGGSGIVONTKALILKSEHRLKEISKLKQNEQLIKQKNELLS 2537
 QY 1613 RVASASLOVGMCMPEPASEREDLVYFEHKGFTKLCKEHDGFTGEDESSAHALERKSDN 1672
 Db 2538 NNQHLNSNEVKTM--KERTLKRE-----AHKQVT--CE-----NS 2567
 QY 1673 PLDIATVTRLADLERNI-----ERRIEDI 1696
 Db 2568 PKSPKVTGTASKKQITPSQCKERNLODPV 2597

RESULT 25
 S62516
 hypothetical coiled-coil protein [imported] - fission yeast (Schizosaccharomyces pombe)
 C:Species: Schizosaccharomyces pombe
 C:Date: 12-Feb-1998 #sequence revision 20-Feb-1998 #text_change 09-Jul-2004
 C:Accession: T38511; T50220; S62516
 R:Jones, L.; Murphy, L.; McNeil, A.; Simpson, I.; Harris, D.; Barrell, B.G.; Rajandream,
 submitted to the EMBL Data Library, October 1995
 A:Reference number: Z21798
 A:Accession: T38511
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1085 <J02>
 A:Cross-references: UNIPROT:Q09863; EMBL:Z66525; NID:G1044926; PIDN:CAA91432.1; PID:G104
 A:Experimental source: strain 972h(-); cosmid c29E6
 R:McDougall, R.C.; Rajandream, M.A.; Barrell, B.G.; Brown, S.; Murphy, L.; Jones, L.; Mc
 submitted to the EMBL Data Library, January 2000
 A:Reference number: Z25046
 A:Accession: T50220
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1085 <MCD>
 A:Cross-references: EMBL:AL136538; PIDN:CAB66473.1; GSPDB:GN00066; SPDB:SPAC30.14c
 A:Experimental source: strain 972h(-); cosmid c30
 C:Genetics:
 A:Gene: SPAC29B6.10c; SPDB:SPAC30.14c
 A:Map position: 1

Query Match 2.9%; Score 296.5; DB 2; Length 1085;
 Best Local Similarity 21.3%; Pred. No. 6.8e-05;
 Matches 196; Conservative 141; Mismatches 270; Indels 315; Gaps 40;

QY 34 KDKKPR-KKAMESNSD--SDSGTSS----- 57
 Db 38 KKKPKAKNVETQANVLLGSAGTDSAVTDVGANNVGDHQAISLELSNLSNRISHPOLV 97
 QY 58 -----DTSS-----EGISSD-----SDLEE-----DEEEDQSIESEDD 90
 Db 98 SANGDDSLINVEGISSTDWQFTLESDDTIEHSPQCILSTGSFVDEAFVPHIPGLEDAT 157
 QY 91 SDSSEAGHKNNQVLLHGIDPRADGQKATEKQAEKIHQPLPLAPESQTHSFSQSQKQ 150
 Db 158 SSNSPNSVQNNMLNTVALGSSSTSRKKKKKKKSKA-----NSASLNVDDQDRDF 208
 QY 151 PQVLSQQLPFIQSSQAKEESVNHKTSVIOQTGLVSNVKNPLSLVNOAKKTYMKLIVPSP 210
 Db 209 EQVSTVDVATYRNGQA-----LSYNGSVRQASMN--VNNK-----SK 248
 QY 211 DVLKAGNKNTSEE-----SSLTTELRSKRQYQKQAFPSQLKQESSKSLKKVIALSNP 265
 Db 249 DIWSSNTEEREQIREFWFLSLSESRSLVKEAVLQKMKEQ----- 293
 QY 266 KATSSSPHAPKQTLNHNHNPFLNALLGNHQPNGVIOEQEAPLATTTTKMQSKINE 325
 Db 294 KYSCSCSVCGKRKL-----AIBEELEVLYDAYEELEQYANTQR 332

QY 326 NIAAASPTPSSPNVNLSTSGRRTPGNTQPPVMPSPASPILHSQCKEKAKAVNNVNPVKTHHS 385
 Db 333 NLANTES-----VNASDEG--SDKSQKGI--SDSP-----KLSIPLNNV----- 369
 QY 386 HPAKSLVEQFPGTSDIPSSKSDSEDSNEDEEDEDDEEDEDDEDDSDSQSESDNSSES 445
 Db 370 -PSKSL-----NDDITQDELNSSNADVDEEVIETTSLEEK 403
 QY 446 DTGSEEBEDDDKQDSDSDTEGEKTSMLKNTKTTSSVKSPSMLTGHSTPRNLHIAKAP 505
 Db 404 NVD-----NQEFVTSISNGQT-----LEDTSHS-----P 428
 QY 506 GSAPALCSESQSAPFLGTSSSTLTSSPHSGTSKRRRVTDRELRIPILEYGHWORETRRN 565
 Db 429 QTQPPF-----QPPY--PSKADEKNVHSD-----LYN 454
 QY 566 FGRL--QGEVAYVAP-----CGKKLROYPEVIKYLSENGIMDISRDNFSAKIRVGD 618
 Db 455 FGSLTVKGGILTVADLLKNVDGKK--FIEMSQLAERRMQREDSNFHPELYESGLE 511
 QY 619 YEARDGPOEMQWCLLKEDVPIPRAMEGRGRPPNPDRQARAEBSRRRRKGRPPNVGN 678
 Db 512 YDEDEEDED-----VDEDELDM-----TDEQRMEEGRMFQ-----1FA 548
 QY 679 AEFLDNADAKLLRKLQA-QETARQAAQIKLRKLQKQEQARVAKAEKQQAIMAAEEKRK 737
 Db 549 ARLEFQ-----RVLQAVREKVAQQAQKLEETEEENKQKQERELKK--IREKEKRD 599
 QY 738 QKEQIKIMKQOEK-----IKRIOQIRMEKELBAQOILEAKKK-----KKEAANAALIEA 787
 Db 600 KKKQKLAKEEERORRAELAEQAQKALSAKQEEARKKREBQRLKREQKKQOELE 659
 QY 788 EKRIKEMERQQAVALLKQHSRERRRRHMLMKAMEARKKAEKERLKOEKDEKRLNKE 847
 Db 560 QKR-EKQKQKEREKELKKQKQOEADREKM-----AREQRLREEEKRILEKRRREKLK 714
 QY 848 RKLQERLELEMAKELKPPNED 869
 Db 715 EE-ERRRREL-----LEKESSE 730

RESULT 26
 A26655
 myosin heavy chain [similarity] - slime mold (Dictyostelium discoideum)
 N:Contains: myosin ATPase (EC 3.6.4.1)
 C:Species: Dictyostelium discoideum
 C:Date: 05-Oct-1988 #sequence revision 05-Oct-1988 #text_change 09-Jul-2004
 C:Accession: A26655; A24728; S00250
 R:Warrick, H.M.; De Lozanne, A.; Leinwand, L.A.; Spudich, J.A.
 Proc. Natl. Acad. Sci. U.S.A. 83, 9433-9437, 1986
 A:Title: Conserved protein domains in a myosin heavy chain gene from Dictyostelium disco
 A:Reference number: A26655; MUID:87092266; PMID:3540939
 A:Accession: A26655
 A:Molecule type: DNA
 A:Residues: 1-2116 <WAR>
 A:Cross-references: UNIPROT:P08799; GB:M14628; GB:M11938; NID:G167834; PIDN:AAA33227.1;
 R:DeLozanne, A.; Lewis, M.; Spudich, J.A.; Leinwand, L.A.
 Proc. Natl. Acad. Sci. U.S.A. 82, 6807-6810, 1985
 A:Reference number: A24728; MUID:86016788; PMID:3901008
 A:Accession: A24728
 A:Molecule type: mRNA
 A:Residues: 2035-2116
 R:Wagle, G.; Noegel, A.; Scheel, J.; Gerisch, G.
 FEBS Lett. 227, 71-75, 1988
 A:Title: Phosphorylation of threonine residues on cloned fragments of the Dictyostelium
 A:Reference number: S00250; MUID:88112226; PMID:2828113
 A:Accession: S00250
 A:Status: nucleic acid sequence not shown
 A:Molecule type: DNA
 A:Residues: 1734-1893 <WAG>
 C:Comment: The rod domain is highly periodic, containing a pattern of 7-residue repeats
 C:Superfamily: myosin heavy chain; myosin motor domain homology
 C:Keywords: actin binding; ATP; coiled coil; hydrolase; nucleotide binding; P-loop; phos

F;1-818/Domain: globular head <HD>
F;89-747/Domain: myosin motor domain homology <MMOT>
F;179-186/Region: nucleotide-binding motif A (P-loop)
F;819-2116/Domain: alpha-helical rod <ROD>

```
Query Match      2.9%; Score 295; DB 2; Length 2116;
Best Local Similarity 18.6%; Pred. No. 0.00017;
Matches 260; Conservative 246; Mismatches 527; Indels 364; Gaps 53;

Qy 14 RKNQEQSQKQPLDARVDKIKKKPRKAMESSNSDSGTSSTSDTSEGISSSDSDDL 73
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 929 RDLSEELQEQKRLNTLEKL-KYEBEELKMGKVNQDQSDTIS--RLEKIKDELQKEVE 985
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 74 E-----DEEEDQIEE-----SEDDSDSEEAQKSNQVNL-----LHGISD 112
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 986 ELTFSFEESKDKGVLETRVRLQSELDLTLRLDSETKDKSELRLQKKLEELKQVQE 1045
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 113 PKA--DQKATEKAQEKRI---HQPFLAFESQTHSFQSQKQPVLSQQLPFFQSSQA 167
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1046 ALAAETAAKLAQEAANKLQGEYTELNEKFNSEVTARSNVEKSKTTLESQ--VAVNNEL 1103
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 168 KEESVNHKT-----SVIQSTGLSVNVKPLSLVNOAKETVMKLIIVSPD 211
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1104 DEEKNRDALFKKKKALDAMLEMKDQLESTG-----GEKSLYDLKVRQES 1150
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 212 VLKAGNNTSESSLLTSELRSKREQVQKAF-----PSQKKQESSKSLKKVI 259
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1151 DMEALRNQISELQSTI-----AKLEKIKSTLEGEVARLQGELEAEQAKSNVEKQKKVVE 1205
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 260 AALSNPKATSSPAHPKQTLNHNPNPLTNALLGNHPNGVISOVQEAFLALTTKTKM 319
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1206 LDLEDKSAQLAEATAAQALDK-----LKKKL---EQELSEVQTQLSEAN----- 1247
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 320 QSKINENIAAASS-----TPFSS-PVNLSTSGRTPGNQTVPMPASPIHL-----SQGK 368
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1248 ---NKNVNSDSTNKHLETSTFNLLKLEAFQAKQALEKRLGLESELKXVNEQLSEEEK 1303
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 369 EKAVNNVNPVKTQHSHSPAKSLVEQFGTDDSDIPSSKQSDSDSNEDE----- 415
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1304 KQKESNEKRVLDSEKVESELKQDIEEVASKAVATEAKNKKESELDSIKQYADVVSERD 1363
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 416 -----BEDDEEDEDDEDDSDSDSSESSESSTEGSEED-----DDDK 458
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1364 KSVQLQTLQAKNEELNTEAEAGQLDRAERSKKKAEFDLEAVKVNLEETAKKVAEK 1423
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 459 DQESDSDTEGETSKMLNKTTSVK-----SPMSLTCHTTPRNHLHTAKAPGS 507
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1424 AMKKAETDYRSTKSELDDAKNVSEQYVQIKRLNEELSELRSVLEEADERCNSAIKAKT 1483
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 508 APAALCESQSPALFTGSSSTLTSSPHSGTSKRRVTDRELRTIPLEVGHORETRIIRFG 567
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1484 AESAL-----ESLKDIEDAANAKAERKSELEVRVA-----ELESLSDKS 1527
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 568 GRLOGEVAYVAPCGKRLQYPEVITKYLRSNGIMDISRDNFSAKIRVGDPFYEARDPQE 627
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1528 GTVNVF-----IRKDAEIDDLRLDRETESI KSDKCKNTRKQFADL-EAKVTEAQ 1581
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 628 MQWCLK-----BEDVI-----PRIRAMEGRGRPPNPDRORAREEEMRRRK 670
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1582 REVVTIDRLKKLESIDIILSTQLDITETKSRITKIEKSKKLEQTLAERRAAEE----- 1634
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 671 GRPNVNAEFLDNADAKLRLQAOETARQAQIKLLRLKLOEQEARVAKEAKQQAIM 730
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1635 -----GSSKAADDE-----IRKQVQWQVDELRAQL-----DSERAAINASEKIKSLV 1677
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 731 AAEEKRKQEQIKIMKQOE--KIKRIQIRMEKELRAQQLLEAKKKKKEEAANAK-----LL 785
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1678 AEVDVEKQLEDEILAKDKLVKAKGALEVELE-EVRDQ--LEEBEDSSELEDSKRLTT 1734
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 786 EAEKRIKEKEMRROQAVLLKHQERERRRQHMLMKAMEARKKAEKERLQKQKDEKRLN 845
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1735 EVEDIKKDYDAEVQNTKLDEAKKKLTDVDTLKKQLE-----DEKKKLNESERAKRLLE 1789
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

```
Qy 846 KEREQRRLLELM-----AKELKKFNEDMCLADQKPLPELPRIPLGLVLSGTSFDCML 899
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1790 SENEDFLAKLDAEVKNSRAEKORKKYEKD--LKDTK----- 1824
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 900 VVOFLRNFQKVLGFDVNDVFNLSVLQEGLLNTGDSMGVEQDILLVRLLSAAVCDPGLITG 959
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1825 -----YKLNDAAATKTQTEIGAAKLEDQIDELRSKLEQSOAKAT-----QA 1865
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 960 YKAKTALGEHLNVGNVNRDNVSEI-LOIFWEAHGCGTETESLKT--KAPQHTPAOKAS 1016
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1866 DKSCKTLEGIDNLRQAQIEDEGKIKMKLEKRALEGELEBELRETVEAEADSKSEASQSK 1925
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1017 VLAFLINELACSKSVSEID-KNI--DYMGNLRDRKVVVEGKRLKRLIIHAKKTGKRDT 1073
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1926 RLVELELEDA-RRLQKEIDAKAEADAKSNLQRE--IVEAKGRL----- 1967
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1074 GGIDLGSEHPLGTPPTGRRKRRKGSDYDDDDDDSDQDDEDEDEDEDEDEDEDEDEDEDE 1133
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1968 -----EESIAARTNSDRSKRL-----EAEIDALTAQVDAEQKAKNQ 2004
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1134 DICEDDEGQOASVELEKQIEKLSKQSOYRRKLFDAHSLSRSVMFGPDPRYRRYWL 1193
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 2005 QI-----KENKTIETELKEYRRK-FGESEKTKTKEF----- 2034
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1194 PROCGIFVEGMESG--EGLEEIAKEREKLLKABSVQIKEEMFETS--GDSLNCNTDHCE 1249
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 2035 -----LVVEKLETDYKRAKKADEQOORLTVEN-DLRKHLSEISLLKDALDKLQRDHDK 2088
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1250 QKBDLKEKONTNLFQK 1266
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 2089 TKRELETETASKIEMQR 2105
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 27
T11583
probable translation initiation factor - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 09-Jul-2004
C:Accession: T11583
R:Barrell, B.G.; Rajandream, M.A.; Walsch, S.V.
submitted to the EMBL Data Library, April 1996
A:Reference number: Z17295
A:Accession: T11583
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1403 <BAR>
A:Cross-references: UNIPROT:Q10475; EMBL:Z73099; NID:g1314152; PID:e241757
A:Experimental source: strain 972h(-)
C:Genetics:
A:Map position: IR
A:Note: SPAC17C9.03

Query Match      2.9%; Score 294; DB 2; Length 1403;
Best Local Similarity 18.8%; Pred. No. 0.00012;
Matches 312; Conservative 221; Mismatches 633; Indels 492; Gaps 65;

Qy 48 NSDSGTSSTSDTSEGISSSDSDDLDEDEEEDEEEOQSEEDDDSDSEEAQKSNQVLL 107
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 2 SSKPPSNTPKFSYARALASSQSN-----KGNSTKASENNTATAEKQAVKPSG-----V 49
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 108 HGISDPKADQKATEKAQEKRIHQPLAFESQTHSFQSQKQPVLSQQLPFFQSSQA 167
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 50 EPTNTSRANAQKTESTGK-----ITSEADETKYNSS-----KSPVN 86
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 168 KEESVNHKTSVIOSTGLSVNVKPLSLVNOAKETVMKLIIVSPDLKAGNKNTSEESLL 227
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 87 KEGSEVKKSEKSGST-----NNKPWR-----GDNTSKPSANS 118
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 228 TSELRSKREYQKAFPSQLKQKQESSKSLKKVIAA---LSNPKATSSPAHPKQTLNHNHP 284
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 119 SAERTS-----SQHQKPTSSQIGKNDNAAPVENVNEKSTSOETAPPVSTV----- 163
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 285 NPFLTALLGNHPNGVISOVQEAFLALTTKTKMQSKINENIAAASSTPFSPVNLSTS 344
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```


Db 1058 SENTSLVESLNAAEVKYANGMIQHSADIQELTRYKAEPFKANDELNOLKSGRESLOQAYD 1117
Qy 832 ERLKQEKDEKRLNKERKLEORRL-----ELE-----MAXELKXP-----N 867
Db 1118 ELLSNAEAKQLDKEREESKRVSDLHALNSNLHDQIEALASKLAVLASQSNPSLN 1177
Qy 868 EDMCLADQAPLPELPRIPGLVLSGTFSD-----CLMVVQFLRNFQKVLGFVNIDV- 919
Db 1178 ESAMGDQ-----SLNASGLTAAEEGRNNEQLLKIKFLRKXKDL--FAAKLDIL 1225
Qy 920 --PMLSVLOE-----GLLNTGDSMGEV-----QDLLVRLLSA-AVCDPGL 956
Db 1226 KAENARLTSEHAIQOKKVDELNGYLNOERAKSOTDVSANKHEEVLARKIETLNAITDSNR 1285
Qy 957 ITGVKAKTALGEHLNLGVNDRNVSEILQIFWEAHCGOTELTE-----SLTKTAP 1006
Db 1286 IL-REERNAL---TLRVAELTDRISSVEKELFPLQCSNKELTSKIEINVENTSLRTEAI 1341
Qy 1007 QAHTPAQKASVLAFLINELACSKSVVSEIDKNIDYMSNLRRDKWVVEGKRLKRIIHAJK 1066
Db 1342 KWR---QANAL-----VEKSNENPEEFKELQAE-----EHLAKL-LTAEKE 1380
Qy 1067 TGKRDTSGGIDLGBEQHPLGPTTGRKRRKGGSDYDDDDDDSDQGBDEBDE--- 1123
Db 1381 LNKQSD---ELTVLKORMNTEI PMLNKQMQLDEARKKQVDEFTNLKQNNTRQTQIME 1437
Qy 1124 -----DKEDQKKTIDICEDEGDQNASVEELEKQIEKLSKOOSQYRRKLPDASHSLR 1177
Db 1438 LKNLLQKEELLKANBELETOKK--TIADKETELQRLKAK----- 1478
Qy 1178 SVMFGPRYRRYRWWILPRCGGIFVEGMSGEGLEIAKEREKLLK-----KAESVQIKEE 1231
Db 1479 -----RYKDFYIGLOSOGG---GTESAELKVRSELEEVNNQLRALKDEHEKITKE 1527
Qy 1232 MFETSGDSLNCSTNDHCEQ-----KEDLKEKONTNLPLOKPGSFKLSK 1275
Db 1528 CDEYKRTPEPTDTSAIRQEKYAKKLKLVLDLTIVARTDLVNQETT--FAGTKSSYDE--- 1582
Qy 1276 LLEVAKMPPESEVMTPKPNAGAN--GCTLSYQNSGKH-----SLGSVQST--ATQS 1322
Db 1583 --TLARLEKELQ-----ENIVANKDINQRLTRENESLHMRINQLTROLGQSQSKPTSS 1635
Qy 1323 NVEKADSNLFPNTGSSGP-----GKFSYSLPFPNDQLLKTLL 1357
Db 1636 VAERG-----NISESPRTANVKPMSSSATVQOSATVTPWRGGE--TPLAS---IRPIS 1684
Qy 1358 EKNQWFSLLP---RTPCDDTSLTHADMSTAS-----LVTPOSQ--- 1393
Db 1685 VQNSRTAAILPTSQQPPAGSSTSTSSSSSSSTSTSSAAGGSSAVAQATALVPPQQQVHT 1744
Qy 1394 -----PPSKS-----PSPTPAPLGSSAQNVPGLNPPFALSPLQVK----- 1427
Db 1745 TGSAALESMASSSTSSHTDYMPTSSASAVAAIIPMGASSAAESIOHPQOND 1804
Qy 1428 ---GG-----VSMMGLOFCGWPTGVVTSNIPFTLSVPSLGSGLGLSE---GNGNSFLT 1474
Db 1805 SOLFVGAAQOQVVALVPRVGGSSSSSTSVPTATAPSIQDGSOSQSPSTSGSSSS 1863
Qy 1475 SNVASSKS-ESPVPQNEKATSAQAAVEYAKVPDFPSPKPIPERMQFGWRIIDPEDLKA 1533
Db 1864 STVVSSHRSRHTPSSSNVTTTQAGCSLOQIKRPRDIEGDSSTGTG----- 1907
Qy 1534 LLKVLHLRGIREKALQKIQKHLDIYITQACLKNKDVAIIELNENEENQVTRDIVENMSVE 1593
Db 1908 -----EGVAEK--MSKITRL-----RGPMSHGELSAGHIGDSGMDVDQMPTSS 1949
Qy 1594 EQAMEMDLSVLQOVEDLERRVASASLOVKGMWCEPASEREDLVYFHKGFTELCKEHDG 1653
Db 1950 QROEDDIQVVDs--DDEEDVLADAD--DG---PIDGGEAEQEGYED-----SYEQDN 1995
Qy 1654 EF---TGEDSSAHALERKSNPLDIAVTRLADLERNIERRIEDIAPLRVWRRALSEA 1710
Db 1996 EMDNNEGGDDNDIAVDAQDNNEVDIEVP-----EHOHQAEESQ-----SLDNQAIATA 2045

Qy 1711 RSAAQV-----ALCICQLOKSIWAKEKSMKVYQCICRKGDNBEILLLCD 1754
Db 2046 SASIQENNQQOATTSGESSENPNVTLPOAEAS-NWQAAASTSTAARRNESS----- 2097
Qy 1755 GCDKGCHTYCHRPKITTPDPGDFPCACIAGSQTLKIKKLHVKGKKTNESKKGKVTL 1814
Db 2098 -----VEIVSSPQVSNFCE-----QPARLESAEVDG--TAEVAGGAPHES 2135
Qy 1815 TGDTEDEDSASTSSSLKR 1832
Db 2136 AGPS-DTGAASVSSPQK 2152
RESULT 33
A54514
glutamic acid-rich protein precursor - malaria parasite (Plasmodium falciparum)
N;Alternate names: GARP
C;Species: Plasmodium falciparum
C;Date: 15-Oct-1994 #sequence_revision 15-Oct-1994 #text_change 09-Jul-2004
C;Accession: A54514
R;Trigilia, T.; Stahl, H.D.; Crewther, P.B.; Silva, A.; Anders, R.F.; Kemp, D.J.
Mol. Biochem. Parasitol. 31, 199-202, 1988
A;Title: Structure of a Plasmodium falciparum gene that encodes a glutamic acid-rich pro
A;Reference number: A54514; MUID:89040048; PMID:2903445
A;Accession: A54514
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-678 <TRI>
A;Cross-references: UNIPROT:P13816; GB:J03998; NID:gl60298; PIDN:AAA29605.1; PID:gl60299
C;Genetics:
A;Introns: 25/3
C;Superfamily: histone H1
C;Keywords: tandem repeat
Query Match 2.8%; Score 288; DB 2; Length 678;
Best Local Similarity 18.0%; Pred. No. 8.7e-05;
Matches 153; Conservative 147; Mismatches 269; Indels 280; Gaps 25;
Qy 374 NNVPVKTQHHSHPAKSLVQFGTSDIPSS-----KQSEDSNEDEEDD-----EE 421
Db 51 NETEKNKDDNSKSETLLKBEKDEKDDVPTTSDNLLKNAHNNNEISSSTDPNTIINVND 110
Qy 422 EDEDEDDSDSDSQSDSDNSDSTEGSEEDDDDKQDSDSDTEGETSMKLNKTTTS 481
Db 111 KDNENSVDKKKD---KKEKKHKDKKEKKKKKKKKKKKKKKKKKKKKKKKK 166
Qy 482 SVKSPSMSTLGHSTPRNLHTAKAPGAPAAPALCSSESQSPAFLTGTSSTLTSSPHSGTSKRR 541
Db 167 NSEVMSLYKTQGHKPKN---ATEHGEENLDEEMVSEINNNAQGGLLSSPY----- 214
Qy 542 RVTDERELRIPLEYGWQRETRIRNFGRLQGEVAYVAPCGKKLRQYDEVTKYLSRNGIMD 601
Db 215 -----QYREQG-----CG-----II-----SSVHE 230
Qy 602 ISRDNFSFSAKIRVGDYFYEARDGPQEMQWCLLKEEDVIPRIAMEGRGRPPNDRQAR 661
Db 231 TSND-----TKNDKE----- 241
Qy 662 EESRMRRRKGRPPNVGNAEPLDNADAKLLRKLQAEIARQAQIKLLRKLQKQEQARVAK 721
Db 242 -----NISDEKKEDHQEEMLTKDKERKQKEKEMQEQEKIEKKKKQBEK 288
Qy 722 EAKKQQAIAAEKRRKQKQEKIKMKQOEKIKRQQIRMEKELRAQQILEAKKKKKEAAN 781
Db 289 EKKQKERRKQKERRKQKEMKQKQKIEKERRKKEE-----KKKKKHDKEN 339
Qy 782 AKLLE-----AEKRIKEKEMRRQOAVLLKH 806
Db 340 EETWQPDQDTSBETNNEIMVPLSPSLTDVTPPEHKEGEHEKHEKGEHKEGHEEH 399
Qy 807 QERERRRQHMLMKAMEARKKAEKERLKOEKDEKRLNKERKLEQRRLLEMAKELKCP 866


```
Qy 1289 MTPKPAGANGCTLSYQNSGKHSLSGVSQSTATQSNVEKADSNLNF----- 1333
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1542 LSLK--AQMNGAS---SRSGKGD--EVESLDIETSLAKINFLNIIADMQQKNDALKAKV 1594
Qy 1334 NTGSSGPKFYSPNPDQLLTKLTENKRNQWFLSLPRTPCD-----DT-----SLTH 1379
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1595 QTLETLPMDFTKPHAFDALTKR-----KPAFLFCDCIDFQDQDHTDCPIQGSSED 1645
Qy 1380 ADMSTASLVTQSQPPSPSP 1401
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1646 QDYSTPSSSNNEKRLPAP 1667

RESULT 35
T30010
hypothetical protein F58G4.1 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C;Accession: T30010
R;Du, Z.; Leimbac, D.
submitted to the EMBL Data Library, February 1996
A;Description: The sequence of C. elegans cosmid F58G4.
A;Reference number: Z20720
A;Accession: T30010
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-1974 <DUZ>
A;Cross-references: UNIPROT:Q21000; EMBL:U05309; PIDN:AAB37057.1; GSPDB:GN00023; CESP:F58G4
A;Experimental source: strain Bristol N2; clone F58G4
C;Genetics:
A;Gene: CESP:F58G4.1
A;Map position: 5
A;Introns: 18/3; 111/3; 164/1; 229/1; 378/1; 440/2; 525/3; 1177/2; 1633/3; 1863/3
C;Superfamily: myosin heavy chain; myosin motor domain homology
F;84-776/Domain: myosin motor domain homology <MMO>

Query Match 2.8%; Score 287.5; DB 2; Length 1974;
Best Local Similarity 19.0%; Pred. No. 0.00033;
Matches 257; Conservative 242; Mismatches 459; Indels 397; Gaps 56;

Qy 18 QEQSKNPDLARVDKIDKPKKAMSSSSSDSDSGTSSDTSSEGISSDSDDEEDDEE 77
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
875 QEERKRDMENARL---EAEKQALLIQLEQERDSSAEGEERSAKLQAKADLEKQWAN 931
Qy 78 BEDQSIESEDDSDSSEAGHKNNQVLLHGISDPKADGQKATEKAQEKRIHQPLPLAF 137
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
932 MNDQLCDEEEKNAALTKQKKIEQDNEGLKTVSDLETTIKQSEKQAKD-HQIRSLQD 990
Qy 138 ESQTH-----SFQSQKQPOVLSQQLPFIQSSQAKESVNHKTSVIOST----- 182
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
991 EIQQQDEVISKLNKEKHQEEVNRKLLIEDIQAESDKVNLNLTAKULESTLDELEDTLER 1050
Qy 183 ---GLVSNVPLSLVNOAKETVMKLIVPSPDVLKAGNKNTSSESSL-----TSELRS 233
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1051 EKRQDCEK-----QRRKVEGELKI---AQELLELNRRHKHEQEQVKKDIELSSIQS 1102
Qy 234 KREYQKQAFPSQLKKQSSKSLKVVIAALSNPKATSSPAHPKQTLNNHPNPLTALL 293
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1103 RLED-EQSLVAKLQKQ-----IKELLARIQELBELDAERNRSRKAEMQMMELEEL 1156
Qy 294 GNH--QPNGVTSVIO-----EAPLALTTKTKMQSKINENIAAASSFPSPVNLSTSGR 346
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1157 GDRUDEAGGATQAGLNLKKEAEILA-----KLKQDLEDA-----INSETS-- 1198
Qy 347 RTPGNQTPVMPSPASPILHSQKKEKAVSNVNPVKTQHSHPAK-----SLVEQPRGTDSD 401
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1199 -----MAALRKINDAVAEISDQDITQKRG----- 1225
Qy 402 IPSSKDSNDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDE 461
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1226 -----KLEREKNDKQREVEDELLQGSADVEAKQRQRCERMAKQLEAQLTMDTLKS 1273
```

```
Qy 462 ESDSDTEGEKTSKMLKNTTSSVKSPSSMLTCHSTPRNLHIKAPGSAALCSBSQSPAF 521
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1274 DEQARLIQELT-MGKNKV-----HNEQDLN--RQLEDAEAQCALNR----- 1313
Qy 522 LGTSSSTLTSSPHSGTSKRRRVTDRELRLPLEYGMQRETRIRNFGRLQGEVAYIAPCG 581
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1314 -----IKQQHSQLEELKRTLQD-----ETRERQ---SLHSQVSNYQLEC 1350
Qy 582 KKLROYPE-----VIKYLNR-NGIMDISRDNFSPSAKIRVGDFFEAR-----DGPOEM 628
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1351 EQPRESLEEBOADKTDVQRLSKANSEIQWRAKEFEGEGVSRABELEETRKLTHKVOEM 1410
Qy 629 QWCLLKEDVIPRIRAMEGRGR-----PNPNDRQARARESMRRRRGRPPNVGNAE 680
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1411 QEQL---ENANQKIGTLEKNKQRLAHDLDAQVDADRANSIASLEKKQKGF-----K 1461
Qy 681 FLNADAK---LIRKQAOEIAQAAQIKLLRLKQKQOARVAKEAKKQQAQIMAAEKRK 737
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1462 VLDEWRRKCEALVAEVSQRETRAAATETFLRNQLEESGEQTEAVKRENKALAQELKD 1521
Qy 738 QKEQI-----KIMKQOEKIKRIQIIRWEKELRAQOILKAKKKKEEANAAIL----- 785
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1522 IADQLGEGKSVHDLQKRR--RLEIEKEBELQALDAECAL--EAEAKVMRAQIEVSQ 1577
Qy 786 ---EAEKRIKEKEMRRQQAVALLKHOERERRRQHMLMKAMEARKKAEEKERLKOERDEK 842
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1578 IRSEIEKRLQEK-----EFPENTRKNHSRTIESMQVSLTESRGR-----AE 1620
Qy 843 RLANKERKLEQRRLLEMAKE-LKKPNEDMCLADQPLPRLPGLVLSGTSFSDCLMVV 901
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1621 LLKTKKLEGDVNELEIADHSNKLNV-----GQKSMKKL-----QDTIREL 1663
Qy 902 QF-----LRNFGKVLGFDVNDVPNLS-----VLOEGLLNIGDSMGVEQDILLVLLSAV 951
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1664 QYQVEEQRSLSE-----SRDHANLAERRSQVLQ-----EKEDL-----AI 1701
Qy 952 CDPGLITGYAKATGALGHLNVGNVNDVSEI-----LQIFMEAHCGQ 994
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1702 YEGSRTFRQAELEAE-----VKDSVNELSNSLLATKRVKVGDLQLL-----Q 1748
Qy 995 TELTESLTKTAFOAHTPAQKASVLAFLINELACSKSVSEIDKNIDYMSNLNRDKWVVEG 1054
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1749 SEIEEAMS---DAKTSDEKAKAIMDASKLA-----DELRSQEHASNLNQSKTTLES 1798
Qy 1055 KLR--KLRIITHAK-----KTGKRTSGGIDLGEBQHPGTPPGKRRRKGGSDSYDDDDD 1108
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1799 QVXDLQMLRDEAEAGIKGKQRLA---KLDMRIHELETELEGENRRH----- 1843
Qy 1109 DSDDDQDEDEDEDEDEKEDQKKTQDICEDEDEGQAASVEELEKQIEKLSKQSQYRRK 1168
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1844 -----ASTQKVLNRKDKRCHELOFQVDEBKKSQ-----ERMYDLLEKLOQKIKTKYRQ 1891
Qy 1169 LPDASHLSRSMVFPDPRYRRYRILPRCGGIFVEGMESGEGLEIAKERELKKAESVQI 1228
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1892 IEDA-----ESLASG-----NLAKYRQLQHVVEDAQE 1918
Qy 1229 KEEMFETSGSLNSCNTDHCQEKEDLKEKDNLF 1263
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1919 RADAEE-----NALQKLKGRSTGSVF 1941

RESULT 36
A39638
plectin - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004
C;Accession: A39638; S21876
R;Wiche, G.; Becker, B.; Luber, K.; Weitzer, G.; Castanon, M.J.; Hauptmann, R.; Stralowa
J. Cell Biol. 114, 83-99, 1991
A;Title: Cloning and sequencing of rat plectin indicates a 466-kD polypeptide chain with
A;Reference number: A39638; MUID:91268156; PMID:2050743
A;Accession: A39638
A;Status: preliminary
```


A:Molecule type: mRNA
A:Residues: 1-4687 <WIC>
A:Cross-references: UNIPROT:P30427; EMBL:X59601; NID:g1292885; PIDN:CAA42169.1; PID:g15158
C:Superfamily: plectin; alpha-actinin actin-binding domain homology; ribosomal protein S6
C:Keywords: cytoskeleton; transmembrane protein
F:6-103/Domain: ribosomal protein S10 homology <RS10>
F:184-399/Domain: alpha-actinin actin-binding domain homology <ACT>

Query Match 2.8%; Score 286; DB 1; Length 4687;
Best Local Similarity 17.8%; Pred. No. 0.0011;
Matches 389; Conservative 359; Mismatches 817; Indels 618; Gaps 86;

Qy 69 SDDLEEDDEEDQSIEP-SDDSDSSESAQHNNQVLLHGISDPKADGCKATEKAQEK 127
Db ||| :
1478 SETLRMEBERLAEQQAERERLAEVAALEKQRQ-LAAEAHAQAQAEARELQ-R 1535

Qy 128 RIHQPLAFESQTHSQSOQKPOVLSQLPTFIQSS-----QAKSESNNKHSTSVIQTG 183
Db :
1536 RMGEVTRREAQVDA-----QQGRSLQEELHQLRQSSEABIQAKAQV----- 1580

Qy 184 LVSNVKPLSLVNQAKKETVMKLIVPSPDVLKAGNKNTSESSLTSELASKREYKQAFP 243
Db :
1581 -----EAEERSRR-TEEEIRVVRLQLETTERQRGGADELQALRARAEAEA 1627

Qy 244 SOLKKQSSKSLLKVIAALNPKNATSSPAHPKOTLENHPNPFLLNALLGNHQPNGVIQ 303
Db :
1628 QKRQAQEAERLARQVQDESQRKQA----- 1653

Qy 304 SVTQEAFLALTTKT-----KMOSKINEN--IAAASSTPPSPYNLST 343
Db ||| ||| :
1654 -----EASLALRVKAAEAAAEKORALQALDELKQAEAEERWLCOAAEARARQVQVALET 1709

Qy 344 SGRRTPGNQTPMVPMSASP-----ILHSQGKEKAVSNVNVPVKTQHHSHPAKS 390
Db :
1710 AQRSAVELQSKRPSFAKTAQLERTLOEBHVTVTLREEARRAQQAERAEAEER 1769

Qy 391 LVQFRGTDSIPSSKDSDNSDNEDEDEDEDEDEDEDEDESDSDSESSTGES 450
Db :
1770 ELERWQ-----LKANEALRLQAEEVAQOQSLAQADAEEKQAEAREARRRKQAEQAV 1824

Qy 451 EEEDDDDKODESDSDTEG-----EKTSMKLNKTTSSVKSPMSLTGHSTPRNLHTA 502
Db :
1825 RORELAPEQLEKORQUTEGTAORLAQAEGLILRLRAETEOGEHORQL-----EEELA 1877

Qy 503 KAPGSAPAALCSSESQSPAFLG---TSSTLTSSPHSGCTSKRRVTDERELRIPLYGWOR 559
Db ||| :
1878 RLQHEATAATQKQLEAEALKVRAEMEVLLASKARAEEESRSTSEKSRQLEAEAGRFR 1937

Qy 560 ETIRNFGGRLOGEVAYYPACGGKKLRQYPEVI-----KYLNRNGIMDISRDNFPSAKI 613
Db ||| :
1938 E-----LABEAARLALABEARHRHELAEBAARQAEADGVLTLEKLAISEATRL 1988

Qy 614 RVGDFFYEARDGPQEWQ-----WCILLKEEDVIPRIAMEG-----RRGRP 653
Db :
1989 KT----EAEIALKEKEAENRLRLAEDEAFORRRUEBQAQHKADIERLAQLRKASES 2044

Qy 654 NPDRQAREESRRRRGRPNV-----GNAE-----FLDNA-DAKLLRLK 693
Db :
2045 ELERQKGLVEDTLRQRQVVEEIIWALKASFEKAAAGKAELELGLRISRNAEDTMWSKEL 2104

Qy 694 QAQEIAPQAQAIKLLRLKQSQARVAKAQQAQIMAAEEKKQKEQIKIMKQQBKIK- 752
Db ||| :
2105 AEQEAARQ-----RLAAEEEEQRREAEERVQSRSLAAEEEAARKV-ALEEVEERLKA 2156

Qy 753 RIQOIRMEKELRAQOILAEAKKKKKEBAANAKLAEKVR-----IKEEMRRQAVLLKHQE 808
Db :
2157 KVPEARLRRE-RAEQESARQLQAQEAARQLQAEEKAHAFVVQQRSEELQOTLQOEQN 2215

Qy 809 REERR-----OHMLMKMAEAKKAEKERLKOEKRDE----- 841
Db :
2216 LERLRSABAARRAAEAEAEAREQAEAAQSRKQVEEAERLKQSAEEQAQAQAQAAA 2275

Qy 842 KRLNKERKLE-QRRLEMAKELKKNPEDCLADQKLPPELPRIPGLVLVSGSTFSCLMW 900

Df 3179 SIYEALRRDILQPEVAVALLEAQAQTGHITDPATSARLTVDEAVRAGLVGPMHEKLLS A 3238

Gy 1811 KVLTLTG--DTEDSDASTSSSLRGNGNDLOQRKMENENTSINLSQSSEFTSVKKPKR-----1864
:
Bd 3239 EKAVTGYRDPSVGSVSFLQALKKG---LIPR--EQGLRL-LDAQLTGTGVDPFSKHVRV 3292

Gy 1865 -----DSKDIALCSMLTETHEDAWPFLPVNKLVPGYKVVIKPMDWDFST 1913
:
Bd 3293 PLDVAYARGYLDRKTNRALT-:::PRDDARVYLDLPST----REPVTYSQ 3334

Gy 1914 IREKLSSGGYPNLTFALDVLRV 1936
:
Bd 3335 LQRCRSDLQLTGLSLPLSEKAV 3357

RESULT 37
S28589 trichohyalin - rabbit
C:Species: Oryctolagus cuniculus (domestic rabbit)
C>Date: 12-Mar-1993 #sequence_revision 01-Mar-1996 #text_change 09-Jul-2004
C:Accession: S28589
R:Fietz, M.J.; Rogers, G.E.
submitted to the EMBL Data Library, December 1992
A:Description: Examination of the gene encoding rabbit trichohyalin.
A:Reference number: S28589
A:Accession: S28589
A:Molecule type: DNA
A:Residues: 1-1407 <PI>
A:Cross-references: UNIPROT:P37709; EMBL:Z19092; NID:g1746; PID:CAA79519.1; PID:g1747
C:Comment: Trichohyalin is a protein of the medulla of the hair and of the inner root sheath.
Covalent modifications to this protein include conversion of arginine to citrulline and C-genetics:

A:Supernotes: 46/3
C:Superfamily: trichohyalin; calmodulin repeat homology
C:Keywords: calcium binding; citrulline; EF hand; hair; tandem repeat
F:49-81/Domain: calmodulin repeat homology <EP2>

Query Match 2.8%; Score 285; DB 1; Length 1407;
Best Local Similarity 19.4%; Pred. No. 0.00028;
Matches 184; Conservative 169; Mismatches 400; Indels 194; Gaps 35;

Gy 3 QTKTSSTSGGNRKCNQESKNQPDLARDVDKI DKKPR --- KKAMESSNSDSGSTSSD 58
:
Bd 98 EKSHGEKGKRLQNRRQDOREFELDRDFEPPERRWKQOERELAESEEQQKRRE 157

Gy 59 TSSEGISSSSD-----DLBEDEEEEOQEIESEDSDSDSEEAQHKSNNQVLLHGI 110
:
Bd 158 RFEQHYSRQRYDKERQRLORQELEERRAAEFQRLRRGRDAEFFIEEQLRRREQ-----211

Gy 111 SDPAKDCOKATEAKAQRIHQIPLAFESO----THISFQSQOQPOVLSQLPFIFQSSQ 166
:
Bd 212 QELKRELREEQQREREQHERALOEBEELQLQRWRREPREQOQLAREIBEIREQ 271

Gy 167 AKESVNKHTSVIQSTGLSNVKPLSLVANCAKETVMKLI VSPDVLKAGNKNTSES SL 226
:
Bd 272 RLSEQERREQLRREQ-----RLUEQERREQUR-----RELEASERERQL 313

Gy 227 LTSSELRSKEQYKOAFPSPOLKYOSSKLKVVTAALSNPKNATSSSPAHPQTTLENHNPNP 286
:
Bd 314 EQSERREQLREBERREQLRKRE-----LEE1-----REREQLREQBERR- 353

Gy 287 FLTNALLGNHPGVQTSVTOEAPLATTTYKMKQSKINENIAASSTPSSPNVLSTS GR 346
:
Bd 354 ----EQLLAE-----VREQAREGESILT--RRWRQLESAGARKQKVSRRPRROEQSL 403

Gy 347 RTPGNOTPVMPSPASPILHSQCKEKANNVNVNPKVTHHHSPAKSLVEQFGRTDSDIPSSK 406
:
Bd 404 RDQERR-----ORDERRELEEQAR-----RQQQWAQEESEERRRQLSAPS LR 449

Gy 407 D-----SDSNSNEDEEDEBEDDEDSDSDSDSNSNESDTEGSEBEDDDD KDQDE 462
:
Bd 450 ERQLRAETEROBORFREEEORRREROELOFEEEELO QRERRAOOLOEEDSFODRER 509

Qy	463	SDSDTE---	GEKTSMLKNKTTSSVKSPSMLTGHSTPRNLHIKA--PGSAPALCSSESQS	518
Db	510	RRRQQEORPQOTWRWOLQE	-----EAQRRRHTLYAKPG	542
Qy	519	PAFIGTSSSTLTSSPHSGTSKRRVTVDERELRIPIEYGVQWORETRINFNGRLOGEVAYYA	578	
Db	543	-----QQEQLREBEELQREKRRQ	-----ERRERYEEBKLOREDEKRR--RRQEREROYR	590
Qy	579	PC-----	GKKLRQVPEYKILSRNGIMDISRDNFSFSAKIRVGD---EYEARDGP	625
Db	591	ELEELRQEEQLRDRKLUREBEQLQEREERLRQERER	-----KLREEEQLLRQESQELR	645
Qy	626	QEWOMCLLKEDVIPR--IRAMEGRGRPPNDPQ	--RAREESMRRRKGRPPNVGNHAFLL	682
Db	646	QERERKLREBEQLLREBEQLRQERERKLUREBEQLLQEREERLRQE--RARKLREBEQL	704	
Qy	683	DNADAKLL-----RKLOAQ	-----TARQAQIKLLRK-----LOKQEOARVAKAKKO	736
Db	705	LRQEEQLRQERERKLUREBEQLLREBEQLLQERDRKLUREBEQLLQSEBEERLRRQEREQ	764	
Qy	727	QAIMAAEERKQEKQIKMKQEQIKIRIQIRMEKELR--AQQILEAKKKK--KEEAANAK	783	
Db	765	QLRRERDRKFRBEQLLQEREERLRQER	-----ERKLUREBEQLLQEREERLRQERERK	821
Qy	784	LLAEAKRIKKEKWRQOAVULLKHQERERR	-----RQHMLMKAMARKKAAEERLKL	835
Db	822	LRBEQLLQEREER	-----LRQERERKLUREBEQLLQEREQLRQERARKLUREBEQLLR	876
Qy	836	QE-----XRDEKRLNKKERLQORRLEMAKELKPPNEDMCLADOK	876	
Db	877	QEEQLRQERDRKLUREBEQLLQEEQLRQERDRKLUREBEQLLQSE	923	

RESULT 38

T16270

hypothetical protein F35D11.11 - Caenorhabditis elegans

C/Species: Caenorhabditis elegans

C/Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 09-Jul-2004

C/Accession: T16270

R/Fulton, B.

submitted to the EMBL Data Library, June 1995

A/Description: The sequence of C. elegans cosmid F35D11.

A/Reference number: Z18487

A/Accession: T16270

A/Status: preliminary; translated from GB/EMBL/DDBJ

A/Molecule type: DNA

A/Residues: 1-1827 <FUP>

A/Cross-references: UNIPROT:Q20042; EMBL:U29381; NID:g868214; PID:g868224; PIDN:AAA68757

A/Experimental source: strain Bristol N2

C/Genetics:

A/Gene: CESP:F35D11.11

A/Introns: 76/2; 131/3; 159/3; 185/3; 221/3; 253/3; 320/1; 869/3; 1133/3; 1205/2; 1250/1;

	Query Match	2.8%;	Score 285;	DB 2;	Length 1827;
	Best Local Similarity	18.9%;	Pred. No. 0.00038;		
	Matches 340;	Conservative 288;	Mismatches 718;	Indels 454;	Gaps 76;

Qy	371	AVGNVNNPVKTHQSHHPAKSLVQFQGTGTDIPSSKSDSDSNDEDEDEDEDEDD	430
Db	181	AVKQQLHELRTTTTNDLDRQLTEFTRCATLMRKAIRHAQKNLDQKE--QMKREKDDVLD	238
Qy	431	ES--DDSQSDSDSNSEDTEGSEEDDDDDKQDE-----SSDSTEGETSKMLNK	478
Db	239	ETURLQNSVTYNTYMKSEERKANERQDLRKKEDECRKLREQNDELSDILSQLSKMAHEMAG	298
Qy	479	TTSSVKSPSM--SLTGHSS---TP-----RNLHIKAPGSAPALCSSESQSPAFGLT	524
Db	299	GRGSSQSPVIVDNTTAHSLFNFTPMDVAKRKLTTK-----NGEIDSEAKAQAE	352
Qy	525	SSSTLTSSPHSGTSKRRVTVDERELRIPIEYGVQWORETRINFNGRLOGEVAYYAPCGKKL	584
Db	353	KERDRAKOLEKEKRRKDDREARRKSSVYS--QREHDLKKLDDLEL-----KASEKI	404

Db 280 AADLFKDQLEAQNVLVVRKVLQEHODEMERENLAHADAIKHRDEELAQAFAELVKVTE 339
QY 251 SSKSLKKVIAALSNPKATSSPAHPKQTLNHNHPFLTNALLGNHPNG----- 300
Db 340 MKMSDVKLVNVEELSELAPA-AAETVRYLRGGQSLSSLVLEHARVRGKLTVEBEDNV 398
QY 301 -----VIOSVQEAFLATTKMQSKNENIAAASSTPFSSPVNLSSTGRTPGNQ 352
Db 399 NLRVTEBELLETIIONKQOMISQKMTDELFDK-----NRFEKQDLAESERRQLLSQ 452
QY 353 TPVMPASPIILHSQGEKAVGNVNPVKTHQSHPAKSLVQFQFGTSDIPSSKDS----- 409
Db 453 RDT-----AQRDLAVRAELEKYQ-RDYEFVSKENAEELY 486
QY 410 -----DNEDEBEDE-----EDEDEDEDESDSQSESDSNSE----- 444
Db 487 AVERQSRMODPNWSEQAQELFONIVQLQRENVELESDIENAKASAAQAINAQSEMAQ 546
QY 445 -----SDTEGSEEDDDKQODESDSDTEGEKTSKLNKTTSSVKSPSMSLTGHSPTNL 499
Db 547 LRADLAVTKSEAEUKTKVEQTKAAFDSLKERTHEFKELVRDSV----- 590
QY 500 HIAKAPGAPAAALCESQSPAFGLTSSSTLTSSPHSGTSKRRVTDRELRIPLEYGWQR 559
Db 591 ---TAAEARTARLRAEEAIAAKVADATIERLRTQAEDYKADHLRREQDL----- 637
QY 560 ETRIRNFGRLOGEVAYVAPCGKKLRQYPEVIKYLNRNGIMDISRD-----FSPSAK 612
Db 638 EQRIRN-----TEANIASVTETNIK-----NAMLDAQKTNTASMQDFKSAK 681
QY 613 IRVGDVEARD-----GQEMQWCLLKEEDVIPRIAMEGRGRPPNPDR 657
Db 682 EKENIFELKVTAVNAENEOURLVDLGRQTLE-AVEQAGSLVRVRSLE-----DEL 732
QY 658 QRAKEESRM-----RRKGRPPNVGNAEFLDNADAKLLKLAQ----- 696
Db 733 QSAATEINLQFANGQBNILEEEOVRMSVVENFVSRVEAERLTHANTQLDVLRLER 792
QY 697 -----ETARQAQIKLRLKQBOARVAKAKQQAQIMAAEKKQKEQIKIMKQOEKI 751
Db 793 DSLKASTRLSDQLTHTKESKLVQORLEKELEIARQLSEKETQVTRDEMELADLRSK 852
QY 752 -----KRIQOIEWEKELRAQOILEAKKKKKEEAAAKLLEAKRIKEK 794
Db 853 ASMSQYTGSDASGWTPLRLKREYQKTRT-QFLESEL-----DDAKRKLLESE--TTQK 905
QY 795 EMRRQQAVALKH-----QERRRRQHMMLKAMEARKKAEKERL--KQEKREKRLNKR 848
Db 906 RMDAEHAISASHNTVLEENLQSQSQMGV-----EKERLVAKAKCFEDRSKQLAE 955
QY 849 KLEQRLELEMAKELKKNEDMCIMQADKPLPELPR---IPGLVLG-----STFSDCLMV 900
Db 956 SLEQNQKLD---ELRSKNDEQLFAHERETNELRRQLQVASLNDGVRRELEVNNNIS 1012
QY 901 VQFLRNFQKVLGPDVNDVPLVLOEG---LJNIGDSMGVQDILLVRL-----LSAA 950
Db 1013 MQ-----NEATRNSALEQHTTIVRFQEDRITIEISANLRLQTELANNKCAAL 1059
QY 951 VCDPGLITGYKAKTALGHH-----LJNVGV-----NRDNVSE-----ILQIFMEAHCGQ 994
Db 1060 VAES---TAKREADQMTHEAERLLQKTEELNSIEENRQQAQYDEKLAQLQIYESLS 1116
QY 995 TELTESLKTAFAQHTPAQKA-----SVLAFL--INELACKSVVSGEIDKNIDYMNLR 1046
Db 1117 ANLNTQNTTMEVKVNTDSSSTVENLQSLQFVRQSKDEATSRAMTAEV-----MRLRL 1171
QY 1047 RDKWVE-GK---LRKLRIIHAUKTG-----KBDTSGGIDLGEEOHPLGTPPGKRKR 1095
Db 1172 AETAIEYERGRNELLRKIRDLTEKATTAALVVEKASLMKEIQALQTDVHNINAKLTEBTK 1231
QY 1096 RKGDSYDDDDDDSDQGEDDEDEDEKQK-----KTDICEDDEGDQ-AA 1146

Db 1232 LOA-----QLHQIQEKADLENQSRSLASNEEQKLKIASSDQEANQKR 1276
QY 1147 SVELEK-----QIEKLSKQSOYRKRLFDASHSLRSV--MFGPDR-----Y 1186
Db 1277 EIBOLKORVOTNARGAASQQLDQKLAQATARQESAAATAKAAEDKFNQTLQAIKY 1336
QY 1187 RRYWIIIPROCGIFVEGMESGEG-----LEEIAKERELK-KAESVOI 1228
Db 1337 RNENTELKK-----LAEAPGEGDPCAARLKLOFDDFTAKINDYKTEIENLNMKVLRMGI 1391
QY 1229 KEEMFETGSLNCNSNDHCEKEDLKEKONTNLF-LQ-----KPG--SPSK----- 1272
Db 1392 LEKSLKNTNDQINQL-----KQENLKLTEINIRMAQLQSVSATDVEKPGPSSASKSVSS 1445
QY 1273 -----LSKLJLEAVAKMPPSEVMT-----PKPNAGANGCTLSYQNSGKHSGLSVQST 1318
Db 1446 IROTPTKVLDPLSAAKQPNEDPTGLTKTSQQPPSPAARPFSGRQAQSSNVIPPAPA 1505
QY 1319 ATOSNVKAD--SNNLFNTGSSGPKFYSPLPDQLLKLTKRQWFLSLPRTPCDDTS 1376
Db 1506 PTPTSQOKVSPVKRPIPPSIPNEPLDIIPVPSDN-IPDTPPTNSFGTVLPVPHHTQTS 1564
QY 1377 L---THADMTASLVTPSQSPKSPSPAPLGSQAQNPVGLNPFALSPLOVKGVSMM 1433
Db 1565 VRVPTQSLFSSSTTTVQPQPEKKNVLPs---IDSAPSTPGNSSMVTSTSSMAPGQSIF 1621
QY 1434 G-LQFCGWPTGVVTSNIPFTLSVP-----SLGSG-----LGLSEGNNS 1471
Db 1622 GNTGNVPVPTTAPTDN---LALPEESVIEGSAQSSLSVSGSIDQRKVQDIDLVDANDGES 1677
QY 1472 FLTSNVASSKS-----ESPVPQNEKATSAQAAPAAVEAKVPDVPFS 1510
Db 1678 RDSINVGGVSSDVRKRTANDFELSEAKRLRESP---NETVTSI----- 1719
QY 1511 PKPIPEEMQFGWRIIDPEDKALLKVLHL---RGIREKALQKIQIQLHLDYITOACLKNK 1567
Db 1720 -ADIFE-----LDDDD--GVLGMEHVSDEDPNDNTIQEORPDVID-----LEND 1761
QY 1568 DVAIIENENEBEQVTRDIVENWSVEQAME---MDLSVLQVEDLERRVASASLQVKG 1623
Db 1762 EEVLDEMDSEEDDDSGFNDEEFDEEEIPEDDDDDDDVVVLSDGDD----- 1807
QY 1624 WMCPEPASEREDLVYPEHKSFTKLCKEHGDEFTGEDESSAHALERKSDNPLDIAVTRLAD 1693
Db 1808 ---EPANDNE-----ESLNDIDDDGIEIEM 1832
QY 1684 LERNIERRIE-----EDIAPLRVRRALSEARSAAQVLCIOQLQKSIaweKSIMKYC 1738
Db 1833 VEESNNRDI EEVLLGGEDSQSLDDQDR---EASAVEAE----- 1869
QY 1739 QICRKGNEBELLLLCDCDCKGCHTYCHRPKITTPDGDWFCPACIAKASGTLKIKLHV 1798
Db 1870 -----DEG-----RDLPGTIDEPS--APADPTGAAG-----IGS 1896
QY 1799 KGKKTNSKKGKVTLLTGDTEDEDSASTSSSLKRGKNDLOKRKMEENTSINLSKQESFTS 1858
Db 1897 SGRMGQVQRVLPTGLRDAEREDQCSSS-----NETNDERPAERLTARNLARMQPTR 1950
QY 1859 VKKPKR 1864
Db 1951 GAKPTR 1956

RESULT 40

T18674

hypothetical protein T04F3.1 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004

C:Accession: T18674; T24464

R:White, S.

submitted to the EMBL Data Library, June 1996

A:Reference number: Z19004

A:Accession: T18674

1577	Db	DKVFGVGTAKKBNDFDEDEKI KRQ-----IASFERSQKEVORSVGAET 1621
751	Qy	IKRIQOIRMEKELRAQOQIILEAKKKKKEEAANAKLLEAEKRIIKEKENRROQAQVALLKHQERE 810
1622	Db	SHSSKHI FDESNI SMDDVNTSQYKXSD---EKLSTPERTVEPVSTATMNL-----1670
811	Qy	RRRQHMLMKAEARKKAEKERLKOEKDEKRLNKKERKLEQRRLEMAKEL---KKPN 867
1671	Db	---DNITFASGIATREK--NTDVLBEERIQKRVEEFKTTTE---NLEIQKVVLTKBEG 1722
868	Qy	EDMCLADQKPLPELPRI PGLVLSGSTFSDCLMVVQLRNFVGKVLGVDNLDVNLVSLVQES 927
1723	Db	DNSDVKDKH-----ASAVNIDLDVFIQRS 1747
928	Qy	GLLINIGDSMEVQDLLVRLLSAAVCDPGLITGYKAKTALGEHLNAGVNRDNVSEILOQIF 987
1748	Db	SKHPENDEDE-----KIRRGIAEF-----1767
988	Qy	MEAHCCQTELTSLTKYKFAQHTPAQKASVLAFLINELACSSVSVSEIDKNIDYMNLR 1046
1768	Db	-----ERTKQEK-----EAQRSTV---IETQYSSKDMFNESIDSLDVWNTSQ 1807
1047	Qy	RDKWVVEGKLRLIHAHKTGRDTSGGIDLGEEOHPLGTPPGKRKRKCGDSYDDDD 1106
1808	Db	-----SDEKLSPEKTVEPVSTATVNLDNMVALSKERRK-----1845
1107	Qy	DDDDSDQGDDEDEDEKDKKKTDCI CEDEDEGDQAASVEELKQIKLSKQO---S 1163
1846	Db	-----ENNETQEEEOIQKRVEEFKESTEE-----QKIQKSIE-LTKECTSD 1887
1164	Qy	QYARKLFDASHLSRVSVMGPDYRRRYRWILPRCGGIFVEGM-----ESGEGLEBIAKE 1216
1888	Db	EKELKTYSGSIDLKV-----FIQGSKKPRNDESDERINRGIAE 1927
1217	Qy	REKLKKAESVOJKEENFETSGDSLNCSDNTHCEQKEDL--KEXDNTNLFQKPGSPSKLS 1275
1928	Db	FERTKQEKBAQ--RSVVVETSPNKHISDESSISMDIEIFRSQDN-----KSTSNFKSGS 1981
1276	Qy	LLEVAKMPPESEVMTPKPNAGANGCTLS-----YQNSGKHIS-----1311
1982	Db	IPIVILPGEKEVASASIN--LNGVFLEKQKQKSTTDYGREKYNKVCENSGSVSTRAMDG 2039
1312	Qy	-----LGSVOSTATQSNVEKADSNLNTGSSGPKFYSPLPNDOLL-----KTLT 1357
2040	Db	ASISLDDIENSTSQKTEKDIDNSOEFPQLSKPVLK--SSISLDDLFNNTSGIEKTTTS 2097
1358	Qy	EKNRQWFSLLPRTPCDDTSLTHADWSTASLV-----TPQS-----QPPSKSPS 1400
2098	Db	EKT-----TTTTTTRTETTTDYTSKRSTSLVDRFGYETATPAISIAAISFPQPSSSSPQ 2153
1401	Qy	PTPAPLIGSSAQNPGVNLNPFALSPLOVKGVSMGLQFCGWPTGVTWTSNIPFTLSVPVSLGS 1460
2154	Db	APFR-----SKQN-----LSLTVPG-----KNWESMMENT-----S 2181
1461	Qy	GLGSLSEGNNSFL---TSNVASSKSESPV-----PQNEKATSAPAAVEVAK 1504
2182	Db	TISLDDSFNNSFSKSNSTQVYEPREKRPLTLPVDNWDNLVSEATNEATKEAPKTPKSDT 2241
1505	Qy	PVDFPSPKPIPEMOPGWRIIDP--EDLKALLKVLHLGIRKALQKQIKH-----1555
2242	Db	LNYSFRPTRISQIKYEW--VADMI GDIDRKNGKGH-----QSEBEHLQKHQDWTWSSV 2293
1556	Qy	-----LDYITQACLKNK-----DVAIIELNEENQ 1581
2294	Db	VYRPTLETSAYTLRSQNLNDCMEIDLNVDFECLMGKKKDDHCECSACRLTEQELBE 2353
1582	Qy	VYRDI V--ENWVSVEQAM-----EMDLSVLQOVEDLERRVASASLQVKGWMCPEPA 1630
2354	Db	IKKGKILELNMWTSQKIIISSEPSDRKSVDFSNPSQI SLNEVFSPEVPLRIETEHAPPS 2413
1631	Qy	S-----ERDLVYFEHKSFTKLCKEHDGEBTGEDSSAHALERKSNPLD---IAVTRLA 1682
2414	Db	TSSPPPESTRYYLSPKVSJETVTTTHOWKDG-----IPMDEIFSPVSSTA 2459


```
Qy 1683 DLERNIERIE-----ED---IAPGLRVRRALSEARSAAQV-----ALCIQQL 1723
Db 2460 DGNRRFSFYEDRSGWDTIGSEDSVMSGGDRGRRS---TRITDQVIDEAFQIFDSQP 2516
Qy 1724 QKSTAWKSIKMYVCQICRKGDNBELLLCDGCDKGCHYCHRPKITIPDGFWFCPACI 1783
Db 2517 STSTAHPKPV-----TETHYDDYVITSLOQED--LDATD 2549
Qy 1784 AKASSQTLKIKKL--HVKGKKTNES-----KKGKVTLTGDTDE 1821
Db 2550 SEVDGENLDVSTFVDDILGKSMDEAFSLSTYSKUREHTDTSIDRKSGEKVHYYNRTD 2609
Qy 1822 DSASTSSSLKGNKDLOKRMEEN-----TSINLSKQESFTSVKPKRDDSDKALCSCM 1875
Db 2610 TSIDKRSKEVITLEDLSELODEIMKLVFVPEVSVKSDSSANIKASQ--NKSTTKPCDE 2667
Qy 1876 ILTMETHOD-----AWPFLLPVNLKLVPGYKVIKK-----PMDFP-----T 1913
Db 2668 ELLEIEIKSEVFLIKGSYLLIPKSDPLGKMLQKLRQNDSLATLDFSLTRKLKLLVNC 2727
Qy 1914 IREKLS-----SGQYPNLET 1928
Db 2728 IREKAGSLDMVMSQYEGLS 2749

RESULT 41
QFHUH
neurofilament triplet H protein - human
N;Alternate names: neurofilament protein, 112K
C;Species: Homo sapiens (man)
C;Date: 30-Jun-1991 #sequence_revision 30-Jun-1991 #text_change 09-Jul-2004
C;Accession: S00979
R;Lees, J.F.; Shneidman, P.S.; Skuntz, S.F.; Carden, M.J.; Lazzarini, R.A.
EMBO J. 7, 1947-1955, 1988
A;Title: The structure and organization of the human heavy neurofilament subunit (NF-H)
A;Reference number: S00979; MUID:88328981; PMID:3138108
A;Accession: S00979
A;Molecule type: DNA
A;Residues: 1-1020 <LEE>
A;Cross-references: UNIPROT:P12036; EMBL:X15306; NID:935028; PIDN:CAA33366.1; PID:g18414
A;Note: it is uncertain whether Met-1 or Met-2 is the initiator
C;Genetics:
A;Gene: GDB:NEFH
A;Cross-references: GDB:120225; OMIM:162230
A;Map position: 22q12.1-22q13.1
A;Introns: 295/1; 361/3; 403/2
C;Superfamily: neurofilament triplet H protein
C;Keywords: coiled coil; heterotrimer; intermediate filament; nerve; phosphoprotein
F;1-100/Domain: amino-terminal <NTD>
F;101-410/Domain: rod #status predicted <ROD>
F;411-1020/Domain: carboxyl-terminal <CTD>
F;502-826/Region: 14-residue repeats
F;503,511,518,526,532,540,546,552,560,566,574,580,586,594,600,606,614,620,628,634,640,644
(covalent) #status predicted
F;732,768/Binding site: phosphate (Thr) (covalent) #status predicted

Query Match 2.8%; Score 284; DB 1; Length 1020;
Best Local Similarity 18.8%; Pred. No. 0.00021;
Matches 188; Conservative 165; Mismatches 366; Indels 282; Gaps 37;

Qy 2 GQTKSTSS---GGNFKC-----NQEOSKNQPLDARVDKIKDKPKKAMESS 46
Db 69 GAASSTDLPTLSNGPEGCMVAVATSRSEKQLQALNDRFAGYIDKVRQLEAHNRSLE-- 126
Qy 47 SNSDSGSGTSDTSSEGISSDSDLEDEDEEDQSS----- 82
Db 127 -----GEAALRQQAGSANGELYREVRNKGAVILGAARGQLRLEQHELLEDIA 179
Qy 83 -IESEDDSDSESAQHKNQVLLHGISDPKA--DGQKATEKAQEK-----RIHQPLP 134
Db 180 HVRQLDDEARQREAEAAA--RALARFAQEAARVDLQKQALQECGGLRRHH--- 234
```

```
Qy 135 LAFESQTHSPSQSQKQPVLSQQLPFIFQSSQAKESVNKHTSVISQSTGLVSNVKPL--S 192
Db 235 -----QEEVCELLGQ-----IQSGAAQQAQMAETRDALKCDVTSALREIRAQ 277
Qy 193 LVNQAKKETWKLIVPSPDVLKAGNKTSESSLLTSELASKRE---QYKQAFPSQLKKQ 249
Db 278 LEHAVQST-----LQSEEFVRVRLDLSEAAKVNNTDAMERSAQBEITEYRQLQARTTEL 332
Qy 250 ESSKSLKVIATAALSNPKATSSPAHPKQTLNHNHPFLTNALLGNHPQNGVISOVQEA 309
Db 333 EALKS-----TKDSLERQSRSELEDHRQADIAS-----YQEAQLQL 367
Qy 310 PLAL--TTKTQWQSKINE-----NIAAA-----SSTPFSFVNLS 342
Db 368 DAELNRTKWEAAQLREYQDLLNVGMALDIEIAAYRKLLEGECECRIGFGPIPFSLPEGL- 426
Qy 343 TSGRRTPGNOTPVWPSASPILHSGGKEKAVNNVNVKTOHHSHPAKSLVQFRTGTSDI 402
Db 427 -----PKIPSVSTHIKVKSEK-----IKVVEKSEKETVIVVE--QTEETQV 466
Qy 403 PSSKSDSDNEDEDEDEDEDEDESDSDSSESDSSESDTEGSESEDEDDDDQDDE 462
Db 467 TEEVTEEEKEAKEEGKEEGEEGEEGEEGEEGEEGEEGEEGEEGEEGEEGEEGEEGEE 523
Qy 463 SDSDEGEKTSMLNKTTTSVKSPSMTLGHSTPRNLHIAPKAGSAPALCSESQSAPFL 522
Db 524 AKSPAFAKSPKEAKSPAFAKSPKE-----AKSPAFAKSPKEAKSPAFAKSPKE 570
Qy 523 GTSSSTLTSSPHSGTSKRRVTDRELRIPLEGWQRETRINFGRLQGEVAYVAPCGK 582
Db 571 EAKSPAFAKSPKEAKSPAFAKSPKE-----EAKSPAFAKSPKEAKSPAFAKSPKE 597
Qy 583 KLQYQEVYKLSRNGIMDISRDNFSESAKIRVGDGYEARDGQEQWQWCLLKEEDVIP-R 641
Db 598 --AKSPAFAKSPKEAKSPAFAKSPKE-----SPKVEAKSPAFAKSPKEAKSPAFAKSPKE 631
Qy 642 IRAMEGRGRPPNPDRORAREEGRMRMR-----KGRPPNVGNAEFLDNADAKLLRLKQ 694
Db 632 VKSPF--KAKSPAFAKSPKEAKSPAFAKSPKEAKSPAFAKSPKEAKSPAFAKSPKE 687
Qy 695 AQIARQAQAIKILRLKQEQARVAKEAK-----KQQAIAAAEKKRQKEQIKI- 744
Db 688 AKGPERAKSPKVEAKSPAFAKSPKEAKSPAFAKSPKEAKSPAFAKSPKEAKSPAFAKSPKE 747
Qy 745 --MKQEKIRIQIRMEKELRAQQL-----EAKKKKKEEANAAL-----LEAK 789
Db 748 EKAKSPAFAKSPKEAKSPAFAKSPKEAKSPAFAKSPKEAKSPAFAKSPKEAKSPAFAKSPKE 807
Qy 790 RIKEKMRQOAVLLKHOBERRRQHMMLMKAMEARKKAEKELKQEKDEKRLNKR- 848
Db 808 KAPEKELPKKEEVKSPKVEAKSPAFAKSPKEAKSPAFAKSPKEAKSPAFAKSPKEAKSPAFAKSPKE 863
Qy 849 --KLEQRRLLEMAKE--LKKPNEDMCLADQKPLPELRIP 885
Db 864 APKEAPKPKVEKKEPAVEKPKESKVEAKKEAEADKKVP 904

RESULT 42
T34036
hypothetical protein B0041.7 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 09-Jul-2004
C;Accession: T34036
R;Fulton, R.; Wohlmann, P.
submitted to the EMBL Data Library, April 1997
A;Description: The sequence of C. elegans cosmid B0041.
A;Reference number: Z21466
A;Accession: T34036
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-1359 <FUL>
A;Cross-references: UNIPROT:Q9U7E0; EMBL:AF000196; PIDN:AAC24256.1; GSPDB:GN00019; CBSP:
A;Experimental source: strain Bristol N2; clone B0041
```


A: Experimental source: strain 129/Ola; testis

C: Genetics:

A: Map position: 10

C: keywords: nucleotidyltransferase

Query Match 2.8%; Score 283.5; DB 2; Length 3122;
Best Local Similarity 17.6%; Pred. No. 0.00082;
Matches 405; Conservative 327; Mismatches 80; Indels 765; Gaps 102;

QY 5 KSTSGGGRK-----CNGEQSK-----QPLDARDVK-----IKDKPRKAMESS 46
DB KSTSGGGRK-----CNGEQSK-----QPLDARDVK-----IKDKPRKAMESS 46
QY 47 SNDSSTGSDTS-----SEGISSDSDDLEED-----EEEB-----DOSI 83
DB SNDSSTGSDTS-----SEGISSDSDDLEED-----EEEB-----DOSI 83
QY 419 AGLESD-GYQGEKRNPLFCHSFGSQNPQSDDEENEPQIEKEEMELSVMSORWDSI 477
DB AGLESD-GYQGEKRNPLFCHSFGSQNPQSDDEENEPQIEKEEMELSVMSORWDSI 477
QY 84 EE-----SEDDSDSEAEQHKSNQVLLHGIDSPKADGOKATEKAQEK 127
DB EE-----SEDDSDSEAEQHKSNQVLLHGIDSPKADGOKATEKAQEK 127
QY 478 EEHCARSLCRNAHRSSTEEDSSSEEMEM-TDNLFLFANLSPIDJG-TADENSDN- 534
DB EEHCARSLCRNAHRSSTEEDSSSEEMEM-TDNLFLFANLSPIDJG-TADENSDN- 534
QY 128 RIHQPLAFE-SOTHS-----FQSQQKQPVLSQOLPPIFOSSQAKESVKNHTS----- 177
DB RIHQPLAFE-SOTHS-----FQSQQKQPVLSQOLPPIFOSSQAKESVKNHTS----- 177
QY 535 -----PLNNENSRASHSVIATSKLSVRPSIFHKAATLEPPSSAKITFECKHTSALSSH 588
DB -----PLNNENSRASHSVIATSKLSVRPSIFHKAATLEPPSSAKITFECKHTSALSSH 588
QY 178 VIQSTGLVSNV-KPLSL-----VNOAKKXTY-MKLIVPSPDVLKAGKN-----TSEESS 225
DB VIQSTGLVSNV-KPLSL-----VNOAKKXTY-MKLIVPSPDVLKAGKN-----TSEESS 225
QY 589 VLNKDGLTDLSPQNSTEGRDNTFTFKESTYSMKYSGSLSTVHSDNSHKEICKDKS 648
DB VLNKDGLTDLSPQNSTEGRDNTFTFKESTYSMKYSGSLSTVHSDNSHKEICKDKS 648
QY 226 LLTSELRSRQYKQAFPSOLKQDESK-SLKVVIALSNPKATSSSPAPKQTLNNH 283
DB LLTSELRSRQYKQAFPSOLKQDESK-SLKVVIALSNPKATSSSPAPKQTLNNH 283
QY 649 LPVSCSESVFYDEEDIPSVTRQVPSRKYSNRKT-----EKDASCIH 691
DB LPVSCSESVFYDEEDIPSVTRQVPSRKYSNRKT-----EKDASCIH 691
QY 284 PNPLFTNALLG-----NHQPNGVQSVIQEAPLALTTKQKQKINENIAAASSTP 334
DB PNPLFTNALLG-----NHQPNGVQSVIQEAPLALTTKQKQKINENIAAASSTP 334
QY 692 VNRHSETILGKNSFNADLH-----SKRKLSSSENEK----- 725
DB VNRHSETILGKNSFNADLH-----SKRKLSSSENEK----- 725
QY 335 FSSPNLSTGRRTPGNTP-----VNPSA-----SPILHSQGEKAVSNVNVNPKTQHSHHP 387
DB FSSPNLSTGRRTPGNTP-----VNPSA-----SPILHSQGEKAVSNVNVNPKTQHSHHP 387
QY 726 -----GNSTLSGVFPFSLTENCOLLPPSGENRMAHLSIESITDESGLNK 770
DB -----GNSTLSGVFPFSLTENCOLLPPSGENRMAHLSIESITDESGLNK 770
QY 388 AKSLVEQFRGTDSIPS-----SKDS----- 408
DB AKSLVEQFRGTDSIPS-----SKDS----- 408
QY 771 LKIRYEFQEHMEKPSLSQAAHYMFPFVSVLNLTRPKLSVTVYKLSQGNKPSRLK 830
DB LKIRYEFQEHMEKPSLSQAAHYMFPFVSVLNLTRPKLSVTVYKLSQGNKPSRLK 830
QY 409 -----EDSNEDEEDDEEDDEEDDEEDDEEDDEEDDEEDDEEDDEEDDEED 457
DB -----EDSNEDEEDDEEDDEEDDEEDDEEDDEEDDEEDDEEDDEEDDEED 457
QY 831 LNKKKLIGLQSTSTKSTGTGATKDSCTHNDLYTGASEKENGLSDSAKATHGTTENKPPT 890
DB LNKKKLIGLQSTSTKSTGTGATKDSCTHNDLYTGASEKENGLSDSAKATHGTTENKPPT 890
QY 458 K-----DQDESDDTEGECT-SMKLNKNTSSVKSPSMSLTGHSTPRNLHIKAPGAPALC 513
DB K-----DQDESDDTEGECT-SMKLNKNTSSVKSPSMSLTGHSTPRNLHIKAPGAPALC 513
QY 891 EHFIDCHPGDGLAEQSGFLYGNKYTLRAK-----RKVNYE----- 927
DB EHFIDCHPGDGLAEQSGFLYGNKYTLRAK-----RKVNYE----- 927
QY 514 SESQSPAFGLTSSSTLTSSPHSGTSKRRVTDRELRIPLEYGWQRETRIRNFGRLQGE 573
DB SESQSPAFGLTSSSTLTSSPHSGTSKRRVTDRELRIPLEYGWQRETRIRNFGRLQGE 573
QY 928 TEDSESEFVTONSK--ISLPH-----PMEIG-----ENLDGTLKSR 961
DB TEDSESEFVTONSK--ISLPH-----PMEIG-----ENLDGTLKSR 961
QY 574 VAYVAPCKKLRQVPEVILKYLNRGIMDISRDNFSFSKIRVGDFYEAH-----DGPQEMQ 629
DB VAYVAPCKKLRQVPEVILKYLNRGIMDISRDNFSFSKIRVGDFYEAH-----DGPQEMQ 629
QY 962 KR--RKMSKKLP--PVIIKYL-----IINFRGRKNMLVKLGIDSKESQVILTBEKELY 1013
DB KR--RKMSKKLP--PVIIKYL-----IINFRGRKNMLVKLGIDSKESQVILTBEKELY 1013
QY 630 WCLIKEDVPIRI-----NAMEGRGRPPNPDRQAREESRMRR-- 668
DB WCLIKEDVPIRI-----NAMEGRGRPPNPDRQAREESRMRR-- 668
QY 1014 KKLAPLKDFWPKVPSDPATKYPILPTPKSHRRSKHSAKKPKGQHRNTSENKIKETL 1073
DB KKLAPLKDFWPKVPSDPATKYPILPTPKSHRRSKHSAKKPKGQHRNTSENKIKETL 1073
QY 669 -----RKGR-----PVN-----VGNAEFLDNADAKLLRKLQAEIARQAQIKLURKLOKQEQ 717
DB -----RKGR-----PVN-----VGNAEFLDNADAKLLRKLQAEIARQAQIKLURKLOKQEQ 717
QY 1074 SFRKKRTHAVLSPSPSPSIAETEDCDLSYSDVMSKLG-----LSERSTS 1118
DB SFRKKRTHAVLSPSPSPSIAETEDCDLSYSDVMSKLG-----LSERSTS 1118
QY 718 RV-----AKEAKQQAIAAEKRRKQEQIKIMKQEKIKRIQOI--RMWELRAQQ 767
DB RV-----AKEAKQQAIAAEKRRKQEQIKIMKQEKIKRIQOI--RMWELRAQQ 767
QY 1119 PINSSPPRCWPTDPRABEIMAAAEKE-----SNLFKPGPNVNTKTVSPRVGKASARA 1172
DB PINSSPPRCWPTDPRABEIMAAAEKE-----SNLFKPGPNVNTKTVSPRVGKASARA 1172

QY 768 ILEAKKKKKEAANAALLAEAKRIKEKEMRRQQAQVLLKHQERRRRRHHMLMKAEARKK 827
DB ILEAKKKKKEAANAALLAEAKRIKEKEMRRQQAQVLLKHQERRRRRHHMLMKAEARKK 827
QY 1173 QV--KKSARLANSSVVT-----NKRNRKNQTTKLVDGKKKPR 1209
DB QV--KKSARLANSSVVT-----NKRNRKNQTTKLVDGKKKPR 1209
QY 828 ABEKERLKQEKREKRL--NKERLQORRLLEL-----EMAKELKKPNEDMCLADOKPL-- 878
DB ABEKERLKQEKREKRL--NKERLQORRLLEL-----EMAKELKKPNEDMCLADOKPL-- 878
QY 1210 AKQORANEKLSKRHAIPADEKMKPHSEALTPNHQSVSELTSSSGAAQALSQKEMSQ 1269
DB AKQORANEKLSKRHAIPADEKMKPHSEALTPNHQSVSELTSSSGAAQALSQKEMSQ 1269
QY 879 -----PELPRIPGLVLSGSTFSDCLMVVQFLRNFQKVLGFDVNIIDVNLVSLQGLLNI 932
DB -----PELPRIPGLVLSGSTFSDCLMVVQFLRNFQKVLGFDVNIIDVNLVSLQGLLNI 932
QY 1270 GPAVDHPLLPQAQGTISAQORLSNC-----FSSLESKSKSVDLRTFPSSRD----- 1315
DB GPAVDHPLLPQAQGTISAQORLSNC-----FSSLESKSKSVDLRTFPSSRD----- 1315
QY 933 GDSNGEVDLLVRLLSAAVCDPGLITGYKATKATLGEHLLNVGVARDNVSEILOIPMEAH 992
DB GDSNGEVDLLVRLLSAAVCDPGLITGYKATKATLGEHLLNVGVARDNVSEILOIPMEAH 992
QY 1316 -DSHSSV-----VYSSGPGISKINIQRSHNSAMFTRKET-- 1350
DB -DSHSSV-----VYSSGPGISKINIQRSHNSAMFTRKET-- 1350
QY 993 GQTELTSLTKTAFOAHTPAQKASVLAFLIN-----ELACSKSVVSEIDKNIDYMSNLR 1047
DB GQTELTSLTKTAFOAHTPAQKASVLAFLIN-----ELACSKSVVSEIDKNIDYMSNLR 1047
QY 1351 -TLIQKSIPLSNHLGVAQSTQVCGIISPKTEESSTOKNCSSMKLNEYRSSL-- 1406
DB -TLIQKSIPLSNHLGVAQSTQVCGIISPKTEESSTOKNCSSMKLNEYRSSL-- 1406
QY 1048 DKWVVEGKRLK--RIIHAKTKGRDTSGGIDLGEHQHPLGTPGRRRRKGGSDSYD 1104
DB DKWVVEGKRLK--RIIHAKTKGRDTSGGIDLGEHQHPLGTPGRRRRKGGSDSYD 1104
QY 1407 -----ESKPEQVCAPNLFHCKDSQQTVS-----VSEQSKTSETCSPGN----- 1445
DB -----ESKPEQVCAPNLFHCKDSQQTVS-----VSEQSKTSETCSPGN----- 1445
QY 1105 DDDDDDDDDDEDEDEDEKEDQKKTIDICEDEDEGDAASVEELEKQIEKLSQKQSQ 1164
DB DDDDDDDDDDEDEDEDEKEDQKKTIDICEDEDEGDAASVEELEKQIEKLSQKQSQ 1164
QY 1446 -----AASEESQTPNCFVTSKAIQAWEQK 1474
DB -----AASEESQTPNCFVTSKAIQAWEQK 1474
QY 1165 YRRKLFASHLSRSMVRGPDYRRY--ILPRC-----GGIIVEGMESGEG 1209
DB YRRKLFASHLSRSMVRGPDYRRY--ILPRC-----GGIIVEGMESGEG 1209
QY 1475 -RGFILDMSN--FPEKVKORSISEAISQTKALSQCKNQNVSTPSVFGESGGLAV 1527
DB -RGFILDMSN--FPEKVKORSISEAISQTKALSQCKNQNVSTPSVFGESGGLAV 1527
QY 1210 LEEIAKEREKLAKEASVQIKEEMFETSGDSLNC--NTDHCQEKEDLKEKDNINFLQKPG 1268
DB LEEIAKEREKLAKEASVQIKEEMFETSGDSLNC--NTDHCQEKEDLKEKDNINFLQKPG 1268
QY 1528 LKELLQKXQ--QKAQSTNVQDSTSTHQPKDNISVSNH--KKANKRTPVTS--PRKP- 1580
DB LKELLQKXQ--QKAQSTNVQDSTSTHQPKDNISVSNH--KKANKRTPVTS--PRKP- 1580
QY 1269 SFSKLSKLELVAKNPPSEVWTP-----KPNAGANGCTLSYONSKHSLGVSQSTAT 1320
DB SFSKLSKLELVAKNPPSEVWTP-----KPNAGANGCTLSYONSKHSLGVSQSTAT 1320
QY 1581 ---RTPRTPKETPRRLKVDPLNLOTSGHLNLSLSDSPILFSDPFCFESYLSLSDLS 1637
DB ---RTPRTPKETPRRLKVDPLNLOTSGHLNLSLSDSPILFSDPFCFESYLSLSDLS 1637
QY 1321 QSNVEKADSNLFWTSGSGPKFYSP--LPNDOLL--KTLTEKNRQWFSLLPRTPCDDT 1375
DB QSNVEKADSNLFWTSGSGPKFYSP--LPNDOLL--KTLTEKNRQWFSLLPRTPCDDT 1375
QY 1638 PEHNYNDINTI---GOTGFCFSYSGQFVPADQNLQKFLUSDVQ---DLFGQAIKDS 1691
DB PEHNYNDINTI---GOTGFCFSYSGQFVPADQNLQKFLUSDVQ---DLFGQAIKDS 1691
QY 1376 SLTHADMTASLAVTPQSPSPSPAPLPGSQAQNPVGLNPPALS----- 1422
DB SLTHADMTASLAVTPQSPSPSPAPLPGSQAQNPVGLNPPALS----- 1422
QY 1692 ELLSHDRQSCS---EEKHVSDSPW---IRASTLNPFLPEKVAAMDNNHRSQWKS 1744
DB ELLSHDRQSCS---EEKHVSDSPW---IRASTLNPFLPEKVAAMDNNHRSQWKS 1744
QY 1423 --PLQVKGVSMMGLQFC-----GWPTG----- 1443
DB --PLQVKGVSMMGLQFC-----GWPTG----- 1443
QY 1745 FHPLTSHSNIME--SFCVQQAENCLTEKSLNRSSVSKEVFLSLPQANSNDWIQGHNRK 1802
DB FHPLTSHSNIME--SFCVQQAENCLTEKSLNRSSVSKEVFLSLPQANSNDWIQGHNRK 1802
QY 1444 ---VVTSNIPFT--LSVP-----SLGSLGLISEGNSFLTSNVASSKSESPVPQNE 1490
DB ---VVTSNIPFT--LSVP-----SLGSLGLISEGNSFLTSNVASSKSESPVPQNE 1490
QY 1803 EADQSLHSANTSFTTILSSPDGELVDAASEDELYVSRNNDVLTLPAPDSSPRSTSSPLOS 1862
DB EADQSLHSANTSFTTILSSPDGELVDAASEDELYVSRNNDVLTLPAPDSSPRSTSSPLOS 1862
QY 1491 KATSAQAAVEVAPVDFP-----SPKPIPEEQFGWNR 1524
DB KATSAQAAVEVAPVDFP-----SPKPIPEEQFGWNR 1524
QY 1863 KNGSFTRTHAILKPLMSPPSREIVATLLDHDLDSEAIYQEPFCNSDVEKPREIGR 1922
DB KNGSFTRTHAILKPLMSPPSREIVATLLDHDLDSEAIYQEPFCNSDVEKPREIGR 1922
QY 1525 IID-----PEDLKALLKVLHRLGIREKALQKIQKHLDYITQ-----ACLKNQDAVITEL 1574
DB IID-----PEDLKALLKVLHRLGIREKALQKIQKHLDYITQ-----ACLKNQDAVITEL 1574
QY 1923 LLAMVETRLPNDLIEFEGDFSLEGL--LWKTA--FSAMTONPRGSPRLNGQAVV-- 1973
DB LLAMVETRLPNDLIEFEGDFSLEGL--LWKTA--FSAMTONPRGSPRLNGQAVV-- 1973
QY 1575 NENPEENQVTRDIVENWVSVEEQAMENDLSVLOQVEDLERRVASASLOVKGWCMCPSPASR- 1633
DB NENPEENQVTRDIVENWVSVEEQAMENDLSVLOQVEDLERRVASASLOVKGWCMCPSPASR- 1633
QY 1974 --NKESNSHNMWED-----KKIIMPCKYAPSRQLVQAMLOAKEEYERS 2016
DB --NKESNSHNMWED-----KKIIMPCKYAPSRQLVQAMLOAKEEYERS 2016
QY 1634 EDLVYFVSHKSTKLCKEHDGEFTGED-----ESSAHALERKSDNPLDI-- 1676
DB EDLVYFVSHKSTKLCKEHDGEFTGED-----ESSAHALERKSDNPLDI-- 1676
QY 2017 KKLPKTELTPVTKSAENVSPSLNPGDCAVSPQVDKCPHTLSSSAHTKEVSKSQIALQT 2076
DB KKLPKTELTPVTKSAENVSPSLNPGDCAVSPQVDKCPHTLSSSAHTKEVSKSQIALQT 2076
QY 1677 ---AVTRLADLERNIERIEED-----IAPGLVRWRRLS-----EAR 1711
DB ---AVTRLADLERNIERIEED-----IAPGLVRWRRLS-----EAR 1711

Db 2077 STTGCSQTLAAASAAVPEDEDDNDNCVSVSYSSPDSGIPPPWQQAASPDFRSLNGDDRH 2136
Qy 1712 SAAQVALC-----IQOLQSIAMKSIKMYCQICRKGNEBELLLCDGCD 1757
Db 2137 SSPGKELCSLAVENFLPKIDGQKSSCSSEWEPQVSIHARARTGKWDPLCL----- 2190
Qy 1758 KGCHT--YCHRPKITTIPOGDFWCFACIAKASGOTLKIKLHVKGKTKNESKKGKVTLT 1815
Db 2191 ---HSTPWQKFLKLEKLEATGLSPLSEP-----KTQKLY--NKGSDADGLRRVULLT 2239
Qy 1816 GDTE-----DEDSASTSSSLKRGNKDLOKQK--MEENTSNLSKQESFTS 1858
Db 2240 TQVENQFAAVNTPKETSQIDGPSLNTYGFKVSIONLQEAALHEIQNLTLISVELHAR 2299
Qy 1859 VKPKRDDSKDALCSM 1875
Db 2300 TRDLQDPDFPICAL 2316

RESULT 45
T29042
hypothetical protein B0228.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T29042
R:Leimbach, D.
submitted to the EMBL Data Library, March 1995
A:Description: The sequence of C. elegans cosmid B0228.
A:Reference number: Z18324
A:Accession: T29042
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-4385 <LEI>
A:Cross-references: UNIPROT:Q8IG62; EMBL:U23168; PIDN:AAC38807.1; CESP:B0228.3
A:Experimental source: strain Bristol N2
C:Genetics:
A:Gene: CESP:B0228.3
A:Introns: 2545/3; 3781/3

Query Match 2.8%; Score 282.5; DB 2; Length 4385;
Best Local Similarity 18.8%; Pred. No. 0.0014;
Matches 386; Conservative 315; Mismatches 702; Indels 653; Gaps 94;

Qy 21 SKNQPLDARDVKDKKPRKAMSS---SNSDS---GTSSSTSEGI-----SSSD-- 68
Db 2586 SSNIQVDSKYDY--RRPAPKNAETFTTESNADAFSLALQEAAGEVTSSTVSNDA 2643

Qy 69 ----SDDL-----BEDEEEDQSTEESEDDSDSESAQK----- 100
Db 2644 KTTVSDKXLSISKTEMKAVSENAVQDHSLLKDSIEAVEMNIPDSRKEILQKNYSIDR 2703

Qy 101 SNNQVLLHG-----ISDPKADGQKATEKAQEKR-----IHQPLP----- 134
Db 2704 SNTVEMQGAISEDIAIQYAPRVNVSQTMRSQKOLLFGGSGFGELEPPLPQEEDE 2763

Qy 135 -LAFESQTHSQSQKQKQVLSQPLPIFOSSQAKEE-----SVNKHTSVI 179
Db 2764 TTAQSRVYRSSSQVRAPSEESIQTALRALSLESLESNARKTFVDRRRENVSLSRKASVE 2823

Qy 180 QSTGLVSNVK-----PLSLVNOAKK--ETVMKLVPSPDVLKAGNKMTSE-----ESSLL 227
Db 2824 RETNWEARVAKADQSIPTVETLQSKKQESIFSSVSVESKDLNVCGSWTTAKPGIAKVSQ 2883

Qy 228 TSE-----LRSKREYQKQAFPSQKKQESSKSLKKVIAALSNP 265
Db 2884 TKKVEKEVTSATMTVASAVECEVGLSKESKESQSDATGSMVR-----AKSIEEVEREFGE 2939

Qy 266 KATSSSPAHPKQTLN---NHPNPFLLT--NALLGNHP--NGVLOSIVQIAPALITTKK 318
Db 2940 VTSKTEKLRDQDQESWIKSIPQRLVEBESAFNGNDEVIIGGVMSL--BAPLEQEEETE 2997

Qy 319 MQSKINENIAAASSTPFPSPVNLSTSGK--RTPGNQTPV-----MPSASFILHSQGEKA 371

Db 2998 KLLTMKSSSEARSLKAATKESIEEADFSEKIEADQELITVQKELVKATGILNA-----A 3052
Qy 372 VSNVNPVKTHHSHHPAKSLVPEQFRGTDSIPSSKOSSEDSNEDEDEDEDEDEDDDE 431
Db 3053 ASREIN-----ASSRMKEY-----SKVPS-----EDV 3073
Qy 432 SDSQSSESDSNSESDTSGSEEDDDDDKQDSESDSTEGEKTSMKL-----NKTTS 482
Db 3074 IELSLTSESRQSNB---GQFKETKEEIVGLWNTGTGENASKVLPHKPPIDTASMKA 3130
Qy 483 VKSPSMLTGHSTPRNLHIAKAPGSAAPALCSE-----SQSPAFLTGTSSTLTSSPHS 535
Db 3131 AKQNSIEWTG-----SLQKSPSAEAMGIVSQKVTGANSKFGIAQAGAVETTLTASAQS 3183
Qy 536 GTSKRRTVDERELRIPLEYQWORETRIRNPGRLQGEVAYYAPCGKKLQYRPIVKYLS 595
Db 3184 GVTG-----RDISVS--NIGVASET--VTQFSNKETG--IGFGA--SNLIAPPPESAETEF 3231
Qy 596 RINGIMDISRDNFSSAKIRVDFYEAR---DGPQEMQWCLLKE---EDVTPTRAMEGRR 649
Db 3232 TGIKISNLSQTSLNKMAASDVGTTVESKIQAPGNYGVDVSLQKVASSDAI--TKMAQASR 3289
Qy 650 GRPPNPDRQAREESRMRRRKGRPPNVGNAEFLD-----NADAKLLRKLQAOAIAR--- 700
Db 3290 DSALISVDFRQRE-----TVSAEKSDLNFKSTNSETQKLKLPESKEEESGIF 3336
Qy 701 -----QAAQIKLLRKLQKQOARVAKEAKQOAIAMAEKR-----KQ 738
Db 3337 IRSSHEYEETQKTLRHRSASRSASRTVTAPTINQEVQMNFDKKVSDSVASGSLSIGVRE 3396
Qy 739 KEQIKMKQOEKIKRTOQIRMEKEL-----RAQQTILEA 771
Db 3397 SSQSEVMQHAERTSETKLSMNEEVAGVRAVSETTNETFRGYOQGVVEVOTGAAMGREA 3456
Qy 772 KKKKKEAANA---KLLAEAKRIKEMERRQAV--LLKHQE---RERRRQHMLMKAM 822
Db 3457 PRQGEAEITQKLRRTLVSERAKASEMAESQTVTQIQKREDSLASEYVRDTLLKSS 3516
Qy 823 EARKKAEK--ERLKQEKDEKLNKERKLEQRLEMAKELKKPNEDMCLADQKPLPE 880
Db 3517 SVSHVATEQMSHVLVMSKSEHISE--KLQERLIEK-----3552
Qy 881 LPRIPGLVLSGSTFSDCLMVVQFL--RNFGKVLGFDYNDVNPNSVLQEGLLNIGDSM--- 936
Db 3553 -----SFSS---QEFISENQGVHTHDV--ID-----NNGDALICW 3583

Qy 937 -----GEVQDILLVLLSAAVCDPLITGYKAKTALGEHLNAGVN 976
Db 3584 KSSETEBQSLDAKQVTEELAGTTLDSLVRLL-----GTDDEKTAEHVI--VG-- 3629

Qy 977 RDNVSEILQIFMEAHCGQTELTESLTKTQPAQHTPAQKASVLAFINELACSKSVVSEID 1036
Db 3630 ---ASEVLNI-----SETRADMEMSRQDAFSDDATVVASDILEE-- 3665

Qy 1037 KNIDYMSNLRDXNVVEGKLRLRIIHAKTGKRDTSGGIDLGEHQPLGTPP-----G 1091
Db 3666 -----RSLSTVH--EFGSEASTTFGVGK-----LVTKKPEKEEVG 3699

Qy 1092 R---KRRRKGGSDDY-----DDDDDDSD-----DQGEDDEDEEDKQKQKKTIDCE 1137
Db 3700 RSPSEYTKLSQFSDITAISETTMDMSEILRLPECSQSVTLISHKQKQKDKDKATV-- 3757

Qy 1138 DEDEGQOASAVELEKQIEKLSKQSQYRRKRLFDASHLSLRSMFGPDVRVRRYYWILPRCG 1197
Db 3758 -ENTADQVTLLEKRE-----VREQSADVKKKPEV--EDIEEVL---PMIRWAEILSMKA 3807

Qy 1198 GIVE--GMSEGEGLIEAKEREKIKAEVSQIK-----EEMF-----ETSG--DS 1239
Db 3808 STSVETQAEINKLSKPEAQASQKYLKANTANLEINKINVNATHEEVMGATTALECKTSGDV 3867

Qy 1240 LNCSDNTHCEKEDLKEKNTNLFQKPGSFKSLKLE-----VAKMPPSEVMTKPFNA 1295

Db 3868 ASIRLDKSRVERVEKYYQENQWLLSTNAEWETLLNDLEDSTVIAQSVQDSWAFSVKASA 3927
Qy 1296 GANGCTLSYQNGKSHLSGVOSTATQSNVEKADSNLFTGSGGPKFYSLPNDQLKLT 1355
Db 3928 TVNLNSDTSIKKAQPEIG-IQKISQSNVEKT-----VGQFSSATIGNELLVT 3974
Qy 1356 LTEKRWQFSLPRTPCDTSLTHA-----DMSTASLVTPOQSPKSPSPPTPA 1404
Db 3975 RLGMDBEIEITL-----VDEINREQAVGGRIRREFGKSETGGGIYLVRRALPKVKETSTHTV 4030
Qy 1405 PLGSSAQNPVGLNPPALPQVKGGVSMGLQFCGWPTGVVTSNIPFTLSVPSLGSGL-- 1462
Db 4031 TVATSFQ-----IFSTMSAGDEIS-----EAVVELIIPSTSVGAEIKS 4069
Qy 1463 GLSENGNSFUTSNVASKSPSPVQ--NEKATSAQPAAVEVAKVPDPPSPKPIPEMQF 1520
Db 4070 TVARSDMTFTSTH-ASEYTASTVADYARDIAVASTAARKKAIPTERTSQK----- 4120
Qy 1521 GWRRIIDPEDLKALLKVLHLGIREKALQKIQKHLDVITQAC-IKNKDVALLIENENEE 1579
Db 4121 -----LKEVGADGIEILSLWEGIETDLDATTQLVDLLRVKSLQTIIESSEE 4166
Qy 1580 NOVTRDIVENNSVERQAMMDLSVLQOVEDLERRVASLOVKGWMCPEPASEREDLVVF 1639
Db 4167 VERINOYLE--WPEEKGLVIHLINVONKETCERNFAVICSJ-----DTVH- 4210
Qy 1640 EHSFTKLCHEGDGFTGEDESSAHALERKSDNPLDIAVTRLADLERNIERREDIAPG 1699
Db 4211 -----SRKSEIPTYSVKVL-TEKRVLETWRVIESG 4242
Qy 1700 -----LRVWRRLSE-----ARSAQ-VALCTOOLQKSIATW-----EKSIM 1734
Db 4243 DVRFNAVILNHRYSLSKPTLAQETVLRVREVRIAAQPLFISAGDIESDVVWSTHOLEKTFV 4302
Qy 1735 KYVCQI-----CRKGDN 1746
Db 4303 REEDRTIIGARKGEN 4318

RESULT 46
T38077
hypotheical coiled-coil protein - fission yeast (Schizosaccharomyces pombe)
C/Species: Schizosaccharomyces pombe
C/Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
C/Accession: T38077
R/Connor, R.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Walsby, S.V.
submitted to the EMBL Data Library, April 1996
A/Reference number: Z21767
A/Accession: T38077
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-1957 <CON>
A/Cross-references: UNIPROT:Q10411; EMBL:Z70690; PIDN:CAA94624.1; GSPDB:GN00066; SPDB:SE
A/Experimental source: strain 972h-; cosmid c1f3
C/Genetics:
A/Gene: SPDB:SPAC1f3.06c
A/Map position: 1

Query Match 2.8%; Score 281.5; DB 2; Length 1957;
Best Local Similarity 18.3%; Pred. No. 0.00057;
Matches 358; Conservative 317; Mismatches 720; Indels 559; Gaps 80;

Qy 18 QEQKNQPLDARVDKIDKKPKKAMESSNSDSGTSSTDSSE-----GISS 66
Db 240 EQASNSLRGEQERLE-----KLLVSNKTVSTLRQTENSIRAECKTLQEKLEKCAINE 294
Qy 67 SDSDDLEE-----DEEEDQSESEDDSDSESEAHQSNQV----- 105
Db 295 EDSKLEELKHNVANSYDAIVHKDKLIEDLSTRISEFPDLKSRDITLSIKNEKLEKLRN 354
Qy 106 LHHGSDPKADQKATE-----KQAEKKIHQPLPLAFESQTHSQSQKQ-----PQ 152
Db 355 TIGSLKDSRTNSQLEEMVELKESNRTIHSQLTDA-ESKLSSEFQENKSLKGSIDEYQN 413

Qy 153 VLSQQLPFIPOSSQAKESVNKHTSVIOSTGLSVNVKPLSLVNOAKKTYMKLIVPSPDV 212
Db 414 NLSSKDKWVKQVSSQLEEA---RSSLAHATGKLAEINSEKDFONKKIKDFEKIEQDLRAC 470
Qy 213 LKAGNVTSESSLL---TSELRSKREQYKQAFPSOLKQKQSSK-SLKKYVAALSNPKAT 268
Db 471 LNSSNELKESALIDKKQDLNNLRQIKR-----QKKVSESTQSSLSLORDILNEK-- 524
Qy 269 SSSPAHPKQTLNNHPNPLTNALGNHQPNGVITQSVQAEAPLALT----- 315
Db 525 ---KKHEVYESQLNELKELGELQTEISNSELSSQLSTLAAEKEAAVATNNELSESNSLQ 581
Qy 316 -----KTKWQSKINENIAAASSTPPSPVNLSTSGRRTPQNG---TPVMPASAPI 362
Db 582 LCNAFOEKLAKSVMLKENEQNFSSLDTSFK---KLNEHQELENHNNHOTITKQLKDTSSK 638
Qy 363 LHS-----QGEKAVSNVNPVKQ---HHSHPAKSLVQFQRTDS-----DI 402
Db 639 LQQLERANPEQKESLTSDENNDLTKLLKLESNSKSLIKQEDVDLSLEKNIOTLKEDL 698
Qy 403 PSSKDS-----EDSNEDEEDDEDEDEDEDDSDSDSDSDSDSDSDSDSDSDSDSDSD 455
Db 699 RKSEALRPFKLEAKNLREVIDNLKGRHETLEAQORNDLHSLSDAKNTNALSSBLTKSS 758
Qy 456 DDKQDSDSDSDEGEKTSMLKNTTSSVKS-PSMSLTGHSTPRNLHIAKAPGSAAPALCS 514
Db 759 EDVKRLTANVETLTQDSKAMKQSFSLVSNYSQISNLVHEL-RDDHV-NMQSQNNLTLES 816
Qy 515 ESQSPAFIGTSSSTLTSSPHSGTSKRRRVTDRELRIPLYGWQRETRIRNFGGRLOQEV 574
Db 817 ESK-----LKTDCENLTQ-----QNTMLIDNVQKLMH--- 843
Qy 575 AYAAPCGKLRQYPEVTKYLSRNGIMDISRDNFSFSAKIRVGDFFYEARDGQEQMWCLLK 634
Db 844 -----KXVQESKVSLEKEVNGKLSLDLKNLSSLNVAISD----- 879
Qy 635 EEDVIPRIAMEGRRGPPNPDRQAREESRMRRKGRPPNVGNABFLDNADAKLRLKQ 694
Db 880 NDQILTQLAELS-----KNVDSLEQESAQLNSGLK 909
Qy 695 AQETARQAQIKLRLKQKQEQARVAKKQQAQIMAAEKKQKEQIKKQKQEKIKRI 754
Db 910 SLEAEKQ-----LLHTENEELHRLDLTKG-----LKTEESKSDGLGKKLTARQSEISNL 960
Qy 755 QOIRMEKELRAQQILEAKKKKEAANAKLLEAE-KRIKEK--EMRROQAVLLKHQER-- 809
Db 961 KEENMS---SQAITSVKSLDETLSKSKLEADIEHLKKNKVSVEVERNALLASNERLM 1017
Qy 810 -----ERRR-----QHMLMKAMEARKKAEKERLK 835
Db 1018 DDLKNGENIASLQTEIEKKGAENDLQSKLSVVSSEVENLLISSQNTKLEDTNQLK 1077
Qy 836 QEKDEKRLKVKERLQORRLELE-----MAKELKKPNEDMCLA 873
Db 1078 YIEKNVQKLDEK--DQRNVELEELTSKYKGLGENAQIKDELLALRKKSK--QHDLCAN 1134
Qy 874 DQKPLPE---LPRIPG---LVLS---GSTFSDCLM---VVQFLRNFGKVLGFDVNI 917
Db 1135 FVDLKEKSDALQTLTEKNEELIVSLEQSNNEALVEERSDLANRLSDMKKLSDSNV 1194
Qy 918 DVPNLVLQELGILNIGDSMGVEQDILLVRLLSAAVCDPLGITGYKAKTALGHLHNVGNR 977
Db 1195 ---ISVIRSDLVRVNDELDTLK-----KDKSLSSTQYSEVCODR 1230
Qy 978 DNVSIEILQIFNEAHCGQOTELTESLTKTFAQHTPAQKASVLAFILN-----ELACSKSV 1032
Db 1231 DDLSDLSKGCESFNKAVSLRELCTKS-EIDVPVSEILDNDNFVNAGNFSEL--SRLTV 1287
Qy 1033 SEIDKNIDYNSLRDKNVVEGKRLKURIHAKTGKEDTSGGIDLGEOHPLGTPTGR 1092
Db 1288 LSLNYLDAPQNVNFKKMLDNRL-----TTDAEFTKVADLEKLOH----- 1330

Qy 1093 KRRKGGSDYDDDDDDDDDDDEDDEDKEDQKKTDTDCEDEDEGDQAASVEELE 1152
Db 1331 -----EHDWLIQRGDEKALKDSEKNFLRKEAEMTE-----NIHSLE 1368
Qy 1153 KQIEKLKQSOYRRKLFDD--ASHSLRSYVMFGPDYRRRYWILPRCGGIFVGMESGEG 1209
Db 1369 EGKEETKKEIAELSSRLDNQLATNKKNLQ---DHUNQEIRL----- 1408
Qy 1210 LEEIAKEREL-----KKAESVQIKKEEMFETSGDSLNCSTNTDHCCEKEDLKE 1256
Db 1409 KEDVLKESKESLIISLESLSNQRKQKESLLDAKNEL-EHMLDTSRKNKSLMEKIESINS 1467
Qy 1257 KONTNLF-----LQKPGSFKLS-----KLLEAVAKPPESEVMTPKPNAGANGCTLSYQNS 1307
Db 1468 SLDDKSPELASAVEKLGALQKLHSELSLMEINIK---SOLQEAKEKIQVDESTIQELD- 1522
Qy 1308 GKHSLSGVQSQTATSNVYKADSNLFTNGSGGPKFYSPLPNDOLLKTLTEKRNQWFSLL 1367
Db 1523 --HEI-----TASKNYE-----GKLD---KDSIIRLSENIEQLNLL 1557
Qy 1368 PRTPC-----DDTSLTHADMSTASLVTPOSPPSK-----SPSPTPAPLGSSAQNP 1413
Db 1558 ABEKSAVKRLSTEKESILQFNRLADLEVHKQVSELSGRSKLKLASTTEELQLAENER 1617
Qy 1414 VGLNPFALS-PLQYKGVSMGLQFCGMPGTGVTSTNIPFTLS--VPSLGSGLGLSEGNG 1470
Db 1618 LSLTTRMLDLQNVKD-----LSNIKDSLSEDLRLRS----- 1650
Qy 1471 SFLTSNVASKSESFPQN-----EKATSAQPAAEVAKPVDVPPSPKPIPEMQFGWNR 1524
Db 1651 --LEDVSASLQKECKISNTVESLDVLTVSQVARNAELEVS-----R 1692
Qy 1525 IIDP-----EDLKALLKVLHLGRREK-----ALQKOIQKHLDYITQACLKNKDVAI 1571
Db 1593 SVDKIRRRDRCEHLGSKLKKLHSQ-LEEQHETTFRAEQQRMTQLGFLKET-VKKOEKLL 1750
Qy 1572 IELNENEENOVTRDIVENWSVEEQAMEDLSVLQQVEDLERRRYASASLQVK-----GWM 1625
Db 1751 KCLNLRQELIPRSSILVYESYIRDIEKEIIVLQE-----RLNGIELSQOLPKGYFGYP 1804
Qy 1626 CPEPASREDLVPEHKSFYTKLCKEHGERT---GEDESSAHALERKSDNPLDIATVRLA 1682
Db 1805 FKTNRVMEVLDGPK-QQVAKLOFAGAEFIVKFKEDLEKCAAEKEKQATFONYSEKVE 1863
Qy 1683 DLERNIER---RIEEDIAPGLRVWRRLSARSAS 1713
Db 1864 NLGKSIEALYFALNREIS-----FRKSLAUSKSA 1892

RESULT 47
T08621
centrosome associated protein CEP250 - human
C;Species: Homo sapiens (man)
C;Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
C;Accession: T08621
R;Mack, G.J.; Reep, J.; Sandblom, O.; Balczon, R.; Fritzler, M.J.; Rattner, J.B.
A;Title: Autoantibodies to a group of centrosomal proteins in human autoimmune sera read
A;Reference number: Z16462; MUID: 98165428; PMID: 9506584
A;Accession: T08621
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-2442 <MAC>
A;Cross-references: EMBL:AF022655; NID:g2832236; PIDN:AAC06349.1; PID:g2832237
A;Experimental source: cell line HeLa

Query Match 2.8%; Score 281.5; DB 2; Length 2442;
Best Local Similarity 17.8%; Pred. No. 0.00074;
Matches 379; Conservative 371; Mismatches 750; Indels 633; Gaps 94;

Qy 11 GGNRKNCKNQESKNQPLDAR-----VDKIKDKPKPKKAMSSNSDSSTGSSD 58
Db 614 GLNQQLQLLEENQSVCSRMEEAAEQARNALQVDLAEAEKREALWEKNTHLLEAQLQKAE 673

Qy	59	TSSEGISSSDDLEDEBEEDQSIBES---EDDDSDSESQAQH---KSNQVILLHGISDP	113
Db	674	AGAE--LQADLRDIOBEKEIQKLSERHQEAAATTOLEQLHQEAKROBEVILAAVQEK	731
Qy	114	KADQKATEKAQEKRIHQPLPLAFESQTHSFQSQOKQOPVLSOQLPFIPOSSQAKESV-	172
Db	732	EA---LVREKA---ALEVR---LQAVERDRODLAAQLOGLSAKELLSSLP	774
Qy	173	--NKHTSVIOSTGLVSNVKPLSLVNQAKKETVMKLIVPSPDVLKAGNQNTSESSLTSE	230
Db	775	EAQQONSVIDEP-----QGOLEVOIQTVTOAKEVIOG-----EVRCLKLE	814
Qy	231	L---RSKREQYKQAFPSOLKQBESSKSLKVIAALSNPKATSSSPAHPTQTLNNHPNF	287
Db	815	LDTERSQAQERDAAAROLAQAQEBQK-----TALEQQA-----AHEKE-----	854
Qy	288	LTNALLGN-----HOPNGVIOSVIQEAPLALT-----KTKMOSKINENIAAASSTPPS	336
Db	855	VNQLREKWEKERSWHQ-----QELAKALESLEKEWLEMLKQOQTEMEA-----	900
Qy	337	SPVNLSTSGRRTPGNOTPVMPSPASPIHSGQKEKAVSNNVNPVKTOHHSHPAKSLVEQFR	396
Db	901	--TQAQREERTQAE-----SALCQMOLTEKERVSLLETLTQTKELADASQOLERL	952
Qy	397	GTOSDIPSSKDSDS-----NEDEEE-----DD-----EEDDEDES	433
Db	953	QDMKVOKLKEQETGILQTOLEQAQRELKEAARQHRDDLAALQESSILLQDMLOLQK	1011
Qy	434	DSQESDSNSESDEGSSEEDDDDK-----DQDESDDTEGETSKMLN-----	477
Db	1012	VEDLKSQVADDSQRLVQEVOQEKLRTOEYNRIQKELEREKASLTLSLMEKEQRLVL	1071
Qy	478	KTTSSVKSPSMSTGHTSPNHLIAKPG-----SAPALCSE---SQSPALGTSSTL	529
Db	1072	QEADSTRQOBSLALQD-----MQAQOGEQKLSAQMELLRQEVKEADFLAQQAQL	1125
Qy	530	TSSPHSGTSKRRVTVBERELRIPLEYGWQRET-----RIRNFGRLQGEVAYVAPCGK	582
Db	1126	ELEAS-----HIT-EQQLRASL---WAGEAKAAQLHLRLRSTESOLEALAAEQP-GN	1174
Qy	583	KLQYQEVIKYLSRNGIMDISRDNFSAKIRVGDIFYEARDGPQEMQWCLKKEEDVIPRI	642
Db	1175	QAAQAQLASLYS-----ALQOALGVSCESR--PE-----LSGGDGAQSPV	1213
Qy	643	RAMEGRRGPPNDRQARAESESRMRERKRPNNVGNAEFLDNADAKLLRKL-----	698
Db	1214	WGLE-----PDQNGARS-----LFRGPLLTALESAEAVASALLKHODLKWTQTRDV	1261
Qy	699	ARQAQIKLLRKL---OKQBOQARVAKBAKQQAQIMAAEKKRQKEQIK-----	743
Db	1262	LRDOVO-KLEERLTDTEAKSQVHTELQDLQRLSQNOBEKSKWECKONSLELMELHE	1320
Qy	744	IMKQEKIKRIQOIRMEKELRAQOILEAKKK---KEEAAVNAKLLAEKR-----	794
Db	1321	TMASLOSRLRFAELRMEQAQ-GERELLQAKENLTAQVEHLQAAVVVEARAQASAAGILSE	1379
Qy	795	EMERQOAVL-LKHQERERRQHMMLMK-----AMAEARKKAEERLKQ	836
Db	1380	DLATARSALKLNEEVESERERAQALQEOGELKVAQKALQENIALLTOTLAEREVEVET	1439
Qy	837	EKDEKRLNKERKLQORRLELEMAKELKXPNEDMCIADQKPIPELPRIPGLVLSGTSFD	896
Db	1440	LRQIQIOELEKQREMQAALDEL-LSLDLKKRNQBVDL-QOEQIOELEK-----	1484
Qy	897	CLMVVOQL-----RNFQKVLGFDVNDIVP-----NLVLOEGLLN-----	938
Db	1485	CRSVLEHLPMAVOERQKLVQREQIRPEKRETOBNVLEHQLLEKXKDWIESQRCQ	1544
Qy	939	VODL-----LVRLLSAV-----CDPLGITGYKA-----KTALGEHLNVGVNRDV	980
Db	1545	VODLKQOLVTLECLALENENHHMKECOOKLKELEGORETORVALTHITLDL-----	1597

RESULT 49

T30336

nuclear/mitotic apparatus protein - African clawed frog

C/Species: Xenopus laevis (African clawed frog)

C/Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 09-Jul-2004

C/Accession: T30336

R/Mendes, A.; Ramyar, K.; Vechio, J.D.; Cleveland, D.W.

Cell 87, 447-458, 1996

A/Title: A complex of NuMA and cytoplasmic dynein is essential for mitotic spindle assembly

A/Reference number: Z20828; MUID:97053784; PMID:8898198

A/Accession: T30336

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: mRNA

A/Residues: 1-2253 <NER>

A/Cross-references: UNIPROT:P70012; EMBL:Y07624; NID:G1514670; PIDN:CAA69905.1; PID:G1514670

C/Genetics:

A/Gene: NuMA

Query Match 2.8%; Score 280.5; DB 2; Length 2253;

Best Local Similarity 19.0%; Pred. No. 0.00074;

Matches 351; Conservative 281; Mismatches 719; Indels 499; Gaps 75;

QY 14 RKNQEQSNQPIQDARV-DKIK--DKPKKAMESSNSDSGTSDDTSSEGIS--S 66

DB 282 KELEELREKNESLMIRLDTLKQCDMDKADKLLERKNDQAEENGELSFKYKVDLSNRLA 341

QY 67 SDSDDLEDEEEDQSTEESEDDDSSES-----EAQHSNNQVLLHG----- 109

DB 342 QLQEALYETTEEQELSLSNNQKQNLQSELSGAVGEEKYLEEHNLTLOQKISWLEDOQLK 401

QY 110 -----ISDPKADGQKATEKAEKRIHQIPLAFESQTHFSQSQOKPOV--LSQQL 158

DB 402 EMEIDMPETGDCMGDILKLDLQKELAVLNTQCLSKQEIHQHMEBEKSTAEVMEAEQKS 461

QY 159 PFIFQSSQAKEESVKNHTSIQSTGLVSNVKPLSLNVQAKKE-----TYMKL----- 205

DB 462 RFSEKQGLQEIVTNLQTSLEIT-----PKERLDNEARAQQEHLWCQITTLKLEISKL 516

QY 206 ---IVSPDVLKAGNKNTSESL-----LTSELRSK-----REQY 238

DB 517 KSSLVHKDEELKGIHHKVEERNEKNQLENFNMGLGNLIGITQQLSEKTKVEDYLREQQ 576

QY 239 KOAF-----PSOLAKQESSKSLKVIASLNPKATSSSPAH-----PKOTLENNH 283

DB 577 QKILCERDSTLSLNEYCKKDNESGVNLTKVTLEQDHTSLSVIEKLKSEKEELASKV 636

QY 284 PNPFNLNALLG-----NHQPNQVIOSVIOEAPLALTTKTKMQSKINENIAAASPTFFSS 337

DB 637 QD--LDKMLGLIAKCNQLNDSQS--KSHAAVTESLKAQLSEQSOLKIYRKVSS 691

QY 338 PVLNLTSGRTPGNQTPMPSASPIHLSQKEK---AVSNVNPVKVTHHSHPAKSLIVEQ 394

DB 692 NELVSENSKLDQLLSVSESLRHLREHLEKTKFAASLADLAKRISHLEEMKKLSE- 750

QY 395 FRGTDSIPSSKDSNEDNEDEEDDEE-----DDEDDSDQSSESDSNSDTEG 449

DB 751 -----SRDEALHNLDERTAGKILSQLKHLSEYQKANESLOAKLAGSCRAIKQ 800

QY 450 SEEDDD-----DKQDQESDSDTEGKTSMLNKTKTSSVYKSPMSLTHSTP 496

DB 801 REERDELSKVVDIWKAKYGESQKIAQNSCHMQEQTEELKKTSDV---YQLEGERSK 857

QY 497 RNLHIAKAPGSAAPALCSSESQSAFLGTSSTLTSPPHSQSTKRRRTVDRELRIPLEYG 556

DB 858 VLMIEAKASSTKSQLEKINQLEGLSAAACT-----KEREAEKKL-VSALHS 906

QY 557 WQSETRIRNPGRLQGEVAYAPCGKKLRQYPEVIKYLNRNGIMDISRDNFESAKIRVG 616

DB 907 AEKLLKI-----AYQGESRSLSHLELTALSNAKQDLCLAKE-LSDEKYKAEFEAMVKV- 959

QY 617 DFYEARDGPOEMQWLLKEEDVPIPRAMEGRGRPPNPDRQAREESRMR-----RRK 670

960	----	-----LKEON-SERTASILESELKNSLAVVYKVERKCESEKLSGEVHLKRO	100	
671	GRPNVGNVAFBFDNADAKLLRLQAOEIA-----	-----ROAAQIKLLRLKLOQBOARVAKE	722	
1003	LDDSSQKHKEALAQNKIETKLINAKEKATSLAIKSEMGALQKQAVDTHKSEFSALQNE	1062		
723	AKQQAIMAABE---KXKQKE---QIKIMKQOEKIKRIQQ-----	-----IRMEKEL	763	
1063	LSRSLDALLAKKEGEVERLNKEAALRQEBEIQOQOQOTITKLTTEETALAALKDKKVALQOEKI	1122		
764	RAQOILLEAKKKKEEAANAK--LLEAEKRIK--EKEMRRQO-----	-----AVLLKHQBERRRRQH	815	
1123	KQO--VOATYGAKEKAVAKLSVISEKSEKIECLEQDIOIQKRDLSLCIOBOHQSKLGEQ	1180		
816	MMLMKA-----MEARKKAEEKERLKQEKRD-----	-----EKLNKERKL	850	
1181	LQALJADLEKKCKEOKELICEAQNKAAEAKTLLASEKASVSERQLEGIOALEIEIGKEROK	1240		
851	E---QRRLELEMAKELKKNEDMCLADQKPLPBLPRI PGLVLSGSTFSD-----	-----CLMVVQF	903	
1241	ACDLOQKQELSWAVOBEKETELQAL--KKEL--FHKVQBELEQSOQTSFTDSSGEGALLYLSE	1296		
904	LRNFGKVL-----GPDVNDVPNLSV--LQEGLLNIGDSNGEVQDILLV-----	-----RLLSA	949	
1297	AQRQOALTEAKBAQEQYQKELEMKNEVNSLQAEIKLSSKVTTNNEVSVDPEQRLKKE	1356		
950	AVCDPGLITGYKAKTALGEBHLNVGNVRD--NVSEILQIFMEAHCGQTE-----	-----	996	
1357	-----TSKSNAK--LBEQWOKLHMELEASFKELEKKNCAIDCLTTEAQNKLGEADQOR	1406		
997	-----LTESLTKVAFQAHTPAKASVL-----	-----AFLINELACSKS	1030	
1407	MAVDSLQQLKSSKAETNHTLQOEBIOAWOKNCAEKEQOICSLQOONLKNSQSLLEEFASLKH	1466		
1031	VYSEIDKNIDYNSLNRDRKMWVEGKRLKLRITHAKTKGRDRTSGGDLGEEQHPGLGTPPT	1090		
1467	SYOEIIAERDLQOEKHQEBELLSHKKLUTE-RFOAELEKAKEDMTETIILLKEKLH-----	-----NQ	1520	
1091	GRKRRKGDSDDYD-----	-----DDDDSDSDQGGDEDD	-----EBDKED	1127
1521	ELQHLKFSQENSYSLTQISHLQOVNSOLLGANGSLSQISDQAKKLESEMSTLKEQHKEE	1580		
1128	QXGKKTTDICEDEDEG-----	-----DQAASVELEK--QIEK	1157	
1581	MKTLRQYKTLREGNKQVQETSLOLETVTYSKYDVKSVKVLKDQKTQBEKORLLLQVQE	1640		
1158	LSKQSOYRKLFPDASHLSRVMFGPDRYRRYRWIILPRCGGIFVEGMESGEGLEETAKER	1217		
1641	LNKQLSQQKBTIRSQOQKUL-----	-----QREGTHEEADKSH	1674	
1218	EKLKKAESYQIKB-----EMFETSGDSLNCSTNDHCEQKEDLKEKONTNLFLOKQPGSF	1272		
1675	KRVLELES-QLEQOTQAVEHYKAQMEKAKVHYDAKKQKQELSEELQSHIKQOE-----	-----H	1728	
1273	LSKLLLEVAKMPPPESEVMTPKPNAGCCTLSYQNSGKLSGLSVQSTATQSNVEKADSNNL	1332		
1729	LSK--ENADLKAESEQL-----	-----HKELQHSL--LQSEKEVQNC	-----NL	1764
1333	FN-----TGS-SGPGKFPYPLPNDQLLTKLTETKNRQWFSLLPRTPCDDTSLTH	1379		
1765	SNRVRSLEASWSTLTQLQDRDPKF-----	-----QLATAFHEESGHFCA--PR-----	-----QTRSH	1810
1380	ADMSTASLVTQSO-----PPSKSPSPPTAPLGSAAQNVPGLNPPFALSPLQVKGVSMMG	1434		
1811	ADVSDLSLSPGEDQLNSTBRCNEEBPATSSVHASSPDSLISGQL-PKQVE---	-----SLES	1866	
1435	LOFCGWPT---GVVTSNIPFTLSVPSLGLGLESGNGNSFLTSNVNASSKSESPVP---	-----	1487	
1867	LYFTPIPTRAQSKLESSIG-SIGDLSLDSKKTQTSARRRTMOIINITWTKTKEEPESAN	1925		
1488	-----QNEKATSAQAAVEVAKPVDFFSPKPIPEMQF	1520		
1926	TSFYSLSRAPSQYSLQHLNPRAGRPAPAIAPALASLPQSQSLAKTEHF	1975		

RESULT 50

T14157

serine/threonine protein kinase - mouse

C;Species: Mus musculus (house mouse)

C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 09-Jul-2004

C;Accession: T14157

R;Pvtowski, B.; Hicklin, D.J.; Kornhaber, G.; Dellaratta, D.V.; Witte, L.

submitted to the EMBL Data Library, December 1997

A;Reference number: Z17894

A;Accession: T14157

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: mRNA

A;Residues: 1-1233 <PVT>

A;Cross-references: UNIPROT:O54988; EMBL:AF039574; NID:g2773155; PID:g2773156; PIDN:AAB5

Query Match

Best Local Similarity 21.3%; Score 280; DB 2; Length 1233;

Matches 208; Conservative 162; Mismatches 393; Indels 214; Gaps 42;

```
Qy 14 RKNQEQ-----SKVQPLDARVDKIDKKPKKAMSSNSDSGTSSTSSGSISSD 68
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 270 RCLKKNVDARWTTSQLQHPFVTVDSGNKPVRELI-----AEAKAE-VTEEV 315
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 69 SDDLDEDEEE-----DOSIESEDD-----DSDSESEAHK 100
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 316 EDGKEDEEEAEANALPANKRASSDLSIASSEDKLSQACILEVSERTEQSTSEDK 375
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 101 SNNQVLLHGISDPKADGQKATEKAQEKRIHQPLAFESQT---HSPSQOQKQPVLSQ 157
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 376 FSNKIL---NEKPTDGPKEAVDEHSDVNLTGAELNDQIVGHIENGREKRPKL--EN 430
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 158 LPFIQSSQ-----AKEESVNGKITSVIOSTGLSVNKPVLNQAQKFTYMKLIVSPD 211
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 431 LPDT-QDQQTVDVNSVSEENNRVTLETNTDC---LKPEDRNKENQETLESKLIQSB 486
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 212 VLKAGNNTSESSLTSELRSREQYKQAPPSQ--LKKOESSKSLAKVIAALSNPKATS 269
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 487 I-----NDTHIQMDLVSQETGEKEADFQAVDNEVGLTQETQKGDGTAKVITSDR 541
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 270 SSPAHPKQTLNHNHPFLTNALGNHPNGVIOSVIQEAPLALTTKTKMQSKINENIAA 329
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 542 SSEVGTDEALDDTKAAELSKAAQSG-----EGDEALVPTQTLAEKPTGPEA 589
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 330 --ASSTFPSPVNLSTSGRTPGNQTPVMPSPASPIHSQKEKAVSNVNPVKTQHSHP 387
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 590 GGAEPEP-----PGGERVEDKQBPQPA---VCEAEGQLTSTSETTRATLEQ----P 634
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 388 AKSLVEQFRGTDS-----IPSSKSDSEDSNEDEDEDEDEDDSDSDSDSDS 441
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 635 ETDEVEQVSESNSIEELRLVVTGAERALGSEGAATAEVDLERKENAQKVPVKAESQA 694
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 442 -----NSEDTEGSEEDDDKQDSDSDSDTEGKTSMLKNTTSS 482
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 695 PAASQSEPHVLPISINSET--TENKEMGALPKETILPPEPEHEKGNDTDSGTGST 753
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 483 VKGPSMLTGHSTPRNLHI-----AKAPGSAPALCSSESQSPAFI-----GTS 525
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 754 VENSSGDL-----NLISISFLSKAKDSGSVLSQETRRQKTKLTKTRKFIVDGEVSVT 806
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 526 SSTLTSSPHSGTSKRR--RYTDERELRIPLEYGQRETRIRNFGRLQGE-----VAYYA 578
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 807 TSKIVTDSKTELRFLRQDELRLK--LQKEEQRAQQQLN--GKLQQOREQIFRRFEQ 863
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 579 PCGKKLRQYPEVIKYLRSNGIMDISRDNFSAKIRVGDYFVARDGPMQWCLLKEEDV 638
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 864 EMLSKKEQYDOEIEENLEKQQTIERLEQETHNELR--DEAKRIKGEQEXE--LSKFQNV 919
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 639 IPRAMEGRGRPPNDRQRAEESRMRRKGRPPNVGNAEFLDNADAKLLRKLQAEI 698
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 920 -----LKNRKKEVMN-EVEKAPRELRLTKRKEBLAQSQ---HAQEQEFVQKQOQEL 969
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
```

```
Qy 699 ARQAAQIKLRLKQKQEQARVAKAE--KQQAIAAAE-----EKRKQKEQIKIMKQEK 750
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 970 ---DGSLLKIIQQKAELANIERECLNNKQOLWRAREAAIWELEERHLQEKHQLKQOLK 1026
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 751 IKRIQIIRMEKELRAQOIIEAKKKKKEEAA--NAKLEAEKRIKEKEMRRQQAVALLKHQE 808
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1027 DQVFMQ-----RHQLKXHEKETBQMQRVYNQRLIEELKNRQTOBRAR----LPKIQR 1074
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 809 RERRRQHMLMKAMEARKKA---EERERLKO-EKRDDEKRLNKERKLSQRRLLEEM----- 859
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1075 SEAKTRMAMPKSLRINSTATPDQREKIKQFAAQEEKQKNERMAHQKHESQMRDLQL 1134
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 860 -----AKELKKPNEDMC 871
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1135 QCEANVRELHQLQNEKC 1151
||| : : : : : : : : : : : : : : : : : : : : : : : : : :
```

Search completed: March 9, 2005, 14:24:02

Job time : 130.833 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 9, 2005, 13:54:12 ; Search time 33.0433 Seconds
(without alignments)
4242.638 Million cell updates/sec

Title: US-10-702-148-13
Perfect score: 9826
Sequence: 1 MEMEANANDHFNLTGLPPA.....MRRFFESRWEEFYOGKQANL 1878

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Minimum Match 0%
Maximum Match 100%
Listing first 65 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/1/iaa/5A.COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B.COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A.COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B.COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCITUS.COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	9816	99.9	1878	4	US-09-418-710-13
2	9816	98.9	1878	4	US-09-839-479-13
3	9712	98.8	1876	4	US-09-418-710-71
4	9712	98.8	1876	4	US-09-839-479-70
5	2482.5	25.3	1972	4	US-09-418-710-21
6	2482.5	25.3	1972	4	US-09-839-479-21
7	2436	24.8	1969	4	US-09-418-710-72
8	2436	24.8	1969	4	US-09-839-479-71
9	1030	10.5	196	4	US-09-690-454-94
10	875.5	8.9	175	4	US-09-418-710-65
11	875.5	8.9	175	4	US-09-839-479-64
12	671.5	6.8	1674	4	US-09-418-710-1
13	671.5	6.8	1674	4	US-09-839-479-1
14	669	6.8	1673	4	US-09-418-710-70
15	669	6.8	1673	4	US-09-839-479-69
16	586	6.0	800	4	US-09-270-767-45282
17	528.5	5.4	1540	4	US-09-949-016-7037
18	521.5	5.3	1525	4	US-09-418-710-69
19	521.5	5.3	1525	4	US-09-839-479-68
20	521.5	5.3	1527	4	US-09-418-710-27
21	521.5	5.3	1527	4	US-09-839-479-27
22	512.5	5.2	1531	4	US-09-418-710-29
23	512.5	5.2	1531	4	US-09-839-479-29
24	472	4.8	176	4	US-09-418-710-66
25	472	4.8	176	4	US-09-839-479-65
26	413.5	4.2	2907	4	US-09-698-295-1
27	406.5	4.1	2781	4	US-09-698-295-10

28	335	3.4	1155	4	US-09-949-016-10125	Sequence 10125, A
29	335	3.4	1155	4	US-09-949-016-10126	Sequence 10126, A
30	308	3.1	59	4	US-09-418-710-52	Sequence 52, Appl
31	308	3.1	59	4	US-09-839-479-51	Sequence 51, Appl
32	288	2.9	2468	4	US-09-976-594-726	Sequence 726, App
33	288	2.9	2468	4	US-09-538-092-1135	Sequence 1135, Ap
34	288	2.9	2522	4	US-09-949-016-10237	Sequence 10237, A
35	273	2.8	45	4	US-09-418-710-62	Sequence 62, Appl
36	273	2.8	45	4	US-09-839-479-61	Sequence 61, Appl
37	272	2.8	3969	4	US-09-538-092-1262	Sequence 1262, Ap
38	260.5	2.7	8991	4	US-08-714-741-32	Sequence 32, Appl
39	257.5	2.6	3969	3	US-08-061-376-5	Sequence 5, Appli
40	256	2.6	52	4	US-09-513-999C-7578	Sequence 7578, Ap
41	252	2.6	1560	4	US-09-264-512B-2	Sequence 2, Appli
42	247.5	2.5	3924	4	US-09-538-092-1246	Sequence 1246, Ap
43	240	2.4	2753	4	US-09-949-016-7659	Sequence 7659, Ap
44	240	2.4	2753	4	US-09-949-016-7660	Sequence 7660, Ap
45	239.5	2.4	1878	4	US-09-949-016-8902	Sequence 8902, Ap
46	239.5	2.4	1878	4	US-09-949-016-8903	Sequence 8903, Ap
47	236.5	2.4	3256	4	US-09-919-172-98	Sequence 98, Appl
48	236.5	2.4	3256	4	US-09-976-594-22	Sequence 22, Appl
49	236.5	2.4	3256	4	US-09-919-039-21	Sequence 21, Appl
50	235	2.4	1043	4	US-09-949-016-11650	Sequence 11650, A
51	233.5	2.4	1219	4	US-09-344-624-4	Sequence 4, Appli
52	229.5	2.3	2972	3	US-09-579-181-2	Sequence 2, Appli
53	229.5	2.3	3118	3	US-09-579-181-1	Sequence 1, Appli
54	228.5	2.3	2289	3	US-09-051-019-2	Sequence 2, Appli
55	227	2.3	2843	1	US-07-741-940-2	Sequence 2, Appli
56	227	2.3	2843	1	US-08-289-548A-2	Sequence 2, Appli
57	227	2.3	2843	1	US-08-452-654-2	Sequence 2, Appli
58	227	2.3	2843	2	US-08-370-235A-2	Sequence 2, Appli
59	227	2.3	2843	4	US-08-449-731-2	Sequence 2, Appli
60	225.5	2.3	1231	4	US-08-714-741-41	Sequence 41, Appl
61	224.5	2.3	2442	3	US-09-514-247A-10	Sequence 10, Appl
62	224.5	2.3	2442	4	US-09-538-092-1370	Sequence 1370, Ap
63	223	2.3	2842	1	US-07-741-940-7	Sequence 7, Appli
64	223	2.3	2842	1	US-08-289-548A-7	Sequence 7, Appli
65	223	2.3	2842	1	US-08-452-654-7	Sequence 7, Appli

ALIGNMENTS

RESULT 1
US-09-418-710-13
; Sequence 13, Application US/09418710
; Patent No. 6596482
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042001
; CURRENT APPLICATION NUMBER: US/09/418,710
; CURRENT FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 1878
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(1878)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-09-418-710-13

Query Match 99.9%; Score 9816; DB 4; Length 1878;
Best Local Similarity 100.0%; Pred. No. 0;

; PRIOR FILING DATE: 1997-10-24;
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 1878
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(1878)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-09-839-479-13

Query Match 99.9%; Score 9816; DB 4; Length 1878;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1878; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MEMEANEDHNFNTGLPPAPAGSLKPSGEGLYNGSPMNFPOQCKSLNGDVNYG 60
DB 1 MEMEANEDHNFNTGLPPAPAGSLKPSGEGLYNGSPMNFPOQCKSLNGDVNYG 60
QY 61 LSTVSHTTTSGILNSAPHSSSTSHLHPSVAYDCLWNYSOYPSANPGSNLKDPPILLSQFS 120
DB 61 LSTVSHTTTSGILNSAPHSSSTSHLHPSVAYDCLWNYSOYPSANPGSNLKDPPILLSQFS 120
QY 121 GGQYPLNGILGGRSPSPSHNTNLRAGSKFWANGTHSPMGLNFDNSOELYDSDPPDQNF 180
DB 121 GGQYPLNGILGGRSPSPSHNTNLRAGSKFWANGTHSPMGLNFDNSOELYDSDPPDQNF 180
QY 181 EYCSGIHPDEAAEKEMTSVVAENGTLVCSLEEXQPELXMCYNGSVPSVESHQEV 240
DB 181 EYCSGIHPDEAAEKEMTSVVAENGTLVCSLEEXQPELXMCYNGSVPSVESHQEV 240
QY 241 VLVPPTVSCLDPPSHLDQLEDTPILSEDSLEPNSLAPRVSGLYIGIDDTELNGAED 300
DB 241 VLVPPTVSCLDPPSHLDQLEDTPILSEDSLEPNSLAPRVSGLYIGIDDTELNGAED 300
QY 301 KLPLXDSPVISALDCPSLNNAAPSLADDDSTSTSIIPASPTSPVPLGESVLQDMSFDLN 360
DB 301 KLPLXDSPVISALDCPSLNNAAPSLADDDSTSTSIIPASPTSPVPLGESVLQDMSFDLN 360
QY 361 NGSDAEQEMETQSSDFPPLSTQAPDQSSSTIQLHPATSPAVSTTSPAVSLVSPASP 420
DB 361 NGSDAEQEMETQSSDFPPLSTQAPDQSSSTIQLHPATSPAVSTTSPAVSLVSPASP 420
QY 421 EISPEVCAPAASTVSPAVSVSPASSAVLPAVSLVPLTASVTSPKASPVTSPPAAFP 480
DB 421 EISPEVCAPAASTVSPAVSVSPASSAVLPAVSLVPLTASVTSPKASPVTSPPAAFP 480
QY 481 ASPANKDVSSLETTADVEEITGEGLTAGSGDVMRRIATPEEVRLPLQHGWRREVRIK 540
DB 481 ASPANKDVSSLETTADVEEITGEGLTAGSGDVMRRIATPEEVRLPLQHGWRREVRIK 540
QY 541 KGSHRWQGETWYGYGCGKMKQFPEVIKYLSENVLVHSVRREHFSPPRMPVGDFFPERDT 600
DB 541 KGSHRWQGETWYGYGCGKMKQFPEVIKYLSENVLVHSVRREHFSPPRMPVGDFFPERDT 600
QY 601 PEGLOWVLSABEIPSRIOAITGKGRPRNTEKAKTEVPKVRGRGPPPKVKITELN 660
DB 601 PEGLOWVLSABEIPSRIOAITGKGRPRNTEKAKTEVPKVRGRGPPPKVKITELN 660
QY 661 TDNRPLKLEAQETLNEDKAKIAKSKMKQKQVORGECLTTIQOAKRKRQETKSLKH 720
DB 661 TDNRPLKLEAQETLNEDKAKIAKSKMKQKQVORGECLTTIQOAKRKRQETKSLKH 720
QY 721 KEAKKSKAEKKGKTKQELKEKVRKEKVKMKKEEVTAKAPACKADTKLATQRL 780
DB 721 KEAKKSKAEKKGKTKQELKEKVRKEKVKMKKEEVTAKAPACKADTKLATQRL 780
QY 781 EERQKQKMLBEMKPTEDMCLTDHQPDPFSRVPGLTLPAGAFSDCLTIVEFLHSFGKV 840
DB 781 EERQKQKMLBEMKPTEDMCLTDHQPDPFSRVPGLTLPAGAFSDCLTIVEFLHSFGKV 840

Db 781 EERQKQKMLBEMKPTEDMCLTDHQPDPFSRVPGLTLPAGAFSDCLTIVEFLHSFGKV 840
QY 841 LGFDPADKDVPSLGLVQEGLLCQGDLSGEVQDILLVRLKKAALHDPGFPSPYCSQSLKIGEKV 900
Db 841 LGFDPADKDVPSLGLVQEGLLCQGDLSGEVQDILLVRLKKAALHDPGFPSPYCSQSLKIGEKV 900
QY 901 SEIPLTRDNVSEIILRCFLMAYGVXPALCDRLRTQFPQAOPOOQKAAVLAFPPVHELNGSTL 960
Db 901 SEIPLTRDNVSEIILRCFLMAYGVXPALCDRLRTQFPQAOPOOQKAAVLAFPPVHELNGSTL 960
QY 961 IINEIDKTLBESMSSYRKNKWIVEGRRLRLKTVLAKRTGRSEVEMGRPEECGLRRSSRIM 1020
Db 961 IINEIDKTLBESMSSYRKNKWIVEGRRLRLKTVLAKRTGRSEVEMGRPEECGLRRSSRIM 1020
QY 1021 EETSGMEEEEEEESIAAVPGRRRRDGEVDATASSIPELERQIEKLSKROLPFRKKLLHS 1080
Db 1021 EETSGMEEEEEEESIAAVPGRRRRDGEVDATASSIPELERQIEKLSKROLPFRKKLLHS 1080
QY 1081 SOMLRVSLGQDRYRRRYWVLPYLAGIFVEGTGEGNLVPEEVIKKTETDSLKVAAHASLNPA 1140
Db 1081 SOMLRVSLGQDRYRRRYWVLPYLAGIFVEGTGEGNLVPEEVIKKTETDSLKVAAHASLNPA 1140
QY 1141 LFSMKMELAGSNNTTASSPARARSRLKTKPGFMOPRHEKSPVRGODSEPOQAOLOPEAQL 1200
Db 1141 LFSMKMELAGSNNTTASSPARARSRLKTKPGFMOPRHEKSPVRGODSEPOQAOLOPEAQL 1200
QY 1201 HVPAPQPOLQLOQSHKGFLQEGSPLSLQSOHDLQSASFSLWSLQTSQSHSSLLSSSV 1260
Db 1201 HVPAPQPOLQLOQSHKGFLQEGSPLSLQSOHDLQSASFSLWSLQTSQSHSSLLSSSV 1260
QY 1261 LTPDSSPGKLDPAQSPQPEPEPEDEASSPDLOAFWFNISAQMPNCAAPTTPPLAVSEDO 1320
Db 1261 LTPDSSPGKLDPAQSPQPEPEPEDEASSPDLOAFWFNISAQMPNCAAPTTPPLAVSEDO 1320
QY 1321 TPSQOOLASSKPMNRPSAANPCSPVQFSSTPLAGLAPKRRAGDPGEMQSPGTGLQPKRR 1380
Db 1321 TPSQOOLASSKPMNRPSAANPCSPVQFSSTPLAGLAPKRRAGDPGEMQSPGTGLQPKRR 1380
QY 1381 GRPSPKPFKQEQRYLTQTAQVPPEMCSGWWIPDPEMLDAMLKALHPGRIGREKALHK 1440
Db 1381 GRPSPKPFKQEQRYLTQTAQVPPEMCSGWWIPDPEMLDAMLKALHPGRIGREKALHK 1440
QY 1441 HLNRHDFLOEVCLRPSADPIFEPRQLPAFQEGIMSNSPKETYETDLAVLQWVEELQR 1500
Db 1441 HLNRHDFLOEVCLRPSADPIFEPRQLPAFQEGIMSNSPKETYETDLAVLQWVEELQR 1500
QY 1501 VIMSDLOIRGWTCPSPDSTREDLAYCEHLSQSDITWRGFGREGLAQPKTTNPDLAV 1560
Db 1501 VIMSDLOIRGWTCPSPDSTREDLAYCEHLSQSDITWRGFGREGLAQPKTTNPDLAV 1560
QY 1561 MRLAALQONVRRYLRPLWPTHEVLEKALLSTPNGAPEGTTTTSYEITPRIWRQT 1620
Db 1561 MRLAALQONVRRYLRPLWPTHEVLEKALLSTPNGAPEGTTTTSYEITPRIWRQT 1620
QY 1621 LQRCRAAHVCLCLGHLESIKAWKSNVKTCLVCRKGDNDDEFLLCLDCGDRGCHYCHR 1680
Db 1621 LQRCRAAHVCLCLGHLESIKAWKSNVKTCLVCRKGDNDDEFLLCLDCGDRGCHYCHR 1680
QY 1681 PKMAVPEGDWFTVCLAQVGEFTQKPGPKRQKSGSYLNFSGDGRRRRLVKG 1740
Db 1681 PKMAVPEGDWFTVCLAQVGEFTQKPGPKRQKSGSYLNFSGDGRRRRLVKG 1740
QY 1741 RESPAAGPRYSEERLSPSKRRLSMRNHSDLTFCETILLMESEHDAAWPFXEPVNPRLV 1800
Db 1741 RESPAAGPRYSEERLSPSKRRLSMRNHSDLTFCETILLMESEHDAAWPFXEPVNPRLV 1800
QY 1801 SGYRRIIKNPMDFSTMRELLRGVTSSEEAADALLVFDNCQTFNEDDSEVGKAGHLMR 1860
Db 1801 SGYRRIIKNPMDFSTMRELLRGVTSSEEAADALLVFDNCQTFNEDDSEVGKAGHLMR 1860
QY 1861 RFFESRWEFFYQKQANL 1878
Db 1861 RFFESRWEFFYQKQANL 1878

RESULT 3		
US-09-418-710-71		
; Sequence 71, Application US/09418710		
; Patent No. 6596482		
; GENERAL INFORMATION:		
; APPLICANT: Jones, Michael H.		
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR		
; FILE REFERENCE: 06501-042001		
; CURRENT APPLICATION NUMBER: US/09/418,710		
; CURRENT FILING DATE: 1999-10-15		
; PRIORITY APPLICATION NUMBER: PCT/JP98/01783		
; PRIOR FILING DATE: 1998-04-17		
; PRIORITY APPLICATION NUMBER: JP 9/310027		
; PRIOR FILING DATE: 1997-10-24		
; PRIORITY APPLICATION NUMBER: JP 9/116570		
; PRIOR FILING DATE: 1997-04-18		
; NUMBER OF SEQ ID NOS: 73		
; SOFTWARE: FASTSEQ for Windows Version 4.0		
; SEQ ID NO 71		
; LENGTH: 1876		
; TYPE: PRN		
; ORGANISM: Homo sapiens		
US-09-418-710-71		
Query Match 98.8%; Score 9712; DB 4; Length 1876;		
Best Local Similarity 99.1%; Pred. No. 0;		
Matches 1861; Conservative 1; Mismatches 14; Indels 2; Gaps 2;		
Qy	1	MEMEANEANDHNFTGLPPAPAAAGLKPSPSSGEGLYTNGSPMNPFOQGKSLNGDVNVNG 60
Dd	1	MEMEANEANDHNFNTGLPPAPAAAGLKPSPSSGEGLYTNGSPHNFPOQGKSLNGDVNVNG 60
Qy	61	LSTVSHTTTTSGILNSAPHSSSTSLLHPHSVAVDCLWYYSQYPSANPGSNLKDPPLLQSRS 120
Dd	61	LSTVSHTTTTSGILNSAPHSSSTSLLHPHSVAYDCUWNYSQYPSANPGSNLKDPPLLQSFS 120
Qy	121	GGOYLPLNGILGSRQSPSSPHNTNLRAQSKFWANGTHSPMGLNFDQSOLYDSDPDQNF 180
Dd	121	GGOYLPLNGILGSRQSPSSPHNTNLRAQSKFWANGTHSPMGLNFDQSOLYDSDPDQNF 180
Qy	181	EVCSGIHPDEAAEKEMTSVVAENGTLGVCSLEBLEXQPDLKMCYGNGSVPSVESLHQEVS 240
Dd	181	EVCSGIHPDEAAEKEMTSVVAENGTLGVCSLELEEEOPELMKCYGNGSVPSVESLHQEVS 240
Qy	241	VLPDPDTVSCLDPPSHLPQLDEDTPILGEDSLSEPNLSLAPBPVSGGLYGIDDTMLGAED 300
Dd	241	VLPDPDTVSCLDPPSHLPQLDEDTPILSEDLSLEPNLSLAPBPVSGGLYGIDDTMLGAED 300
Qy	301	KLPLXDSPVISALDCPSLNATAFSLADDSQTSTSIFASPTSPPVLTGESVLQDNSFDLN 360
Dd	301	KLPLEDSPVISALDCPSLNATAFSLADDSQTSTSIFASPTSPPVLTGESVLQDNSFDLN 360
Qy	361	NGSDAQESEMETSDFPSSLTQPADDSSTLIQLHPATSPAVPSTTSPAVSLVVSIPAASP 420
Dd	361	NGSDAQESEMETSDFPSSLTQPADDSSTLIQLHPATSPAVPSTTSPAVSLVVSIPAASP 420
Qy	421	EISPEVCPAASTTVSPAVSVSPASSAVLPVSLVEPLTASVTSPKASPVTSIPAAAAPT 480
Dd	421	EISPEVCPAASTTVSPAVSVSPASSAVLPVSLVEPLTASVTSPKASPVTSIPAAAAPT 480
Qy	481	ASPANKDVSSFLETADVBEEITGEGLTASGGSDVMRRRIATPEEVRLPLQHGWRREVRIR 540
Dd	481	ASPANKDVSSFLETADVBEEITGEGLTASGGSDVMRRRIATPEEVRLPLQHGWRREVRIR 540
Qy	541	KGSHRWQGETTWYGPCGRMKQFPVIVKYLSRNLVHSVVRSHSFSPRMPVGDFFBERDT 600
Dd	541	KGSHRWQGETTWYGPCGRMKQFPVIVKYLSRNLVHSVVRSHSFSPRMPVGDFFBERDT 600
Qy	601	PEGLOWVQLSAEIPSRIOAITGKRGRPRANTEKAKTEVPVKYRGGRGPPPKVKITELLNK 660
Dd	601	PEGLOWVQLSAEIPSRIOAITGKRGRPRANTEKAKTEVPVKYRGGRGPPPKVKITELLNK 660

QY 1741 RESPAAGPRYSBERLSPSKRRLSRNHHSDLTFCBIIIMENESHDAAWPFYEPVNPRLV 1800
Db 1739 RESPAAGPRYSBERLSPSKRRLSRNHHSDLTFCBIIIMENESHDAAWPFLEPNPRLV 1798
QY 1801 SGVRIIKNMPDFTWRELLRGYTSSEPAADALLVFDNCOTNEDDSEVGKAGHMR 1860
Db 1799 SGVRIIKNMPDFTWRELLRGYTSSEPAADALLVFDNCOTNEDDSEVGKAGHMR 1858
QY 1861 REFESRWEFFYQKQANL 1878
Db 1859 REFESRWEFFYQKQANL 1876
RESULT 4
US-09-839-479-70
; Sequence 70, Application US/09839479
; Patent No. 6727222
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042002
; CURRENT APPLICATION NUMBER: US/09/839,479
; CURRENT FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 70
; LENGTH: 1876
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-839-479-70

Query Match 98.8%; Score 9712; DB 4; Length 1876;
Best Local Similarity 99.1%; Pred No. 0;
Matches 1861; Conservative 1; Mismatches 14; Indels 2; Gaps 2;
QY 1 MEMEANEANDHFNFTGLPAPAASGLKPSPPSGEGLYTNGSPMFPQCKSLNGDVNYG 60
Db 1 MEMEANEANDHFNFTGLPAPAASGLKPSPPSGEGLYTNGSPHNPQCKSLNGDVNYG 60
QY 61 LSTVSHTTTSGILNAGPHSSSTSHLHPSVAVDCLWNTSYQPSANPGSNLKDPPLLSQFS 120
Db 61 LSTVSHTTTSGILNAGPHSSSTSHLHPSVAVDCLWNTSYQPSANPGSNLKDPPLLSQFS 120
QY 121 GQGYPLNGILGSGROPSSPHNTNLRAGSKFWANGTHSPMGLNFDSDQLYDSFPDQNF 180
Db 121 GQGYPLNGILGSGROPSSPHNTNLRAGSKFWANGTHSPMGLNFDSDQLYDSFPDQNF 180
QY 181 EYCSGHPDEAAEKEMTSVVAENGTLVCSLEBEXQPELKMCGYNGSVPSVSLHQEVS 240
Db 181 EYCSGHPDEAAEKEMTSVVAENGTLVCSLEBEXQPELKMCGYNGSVPSVSLHQEVS 240
QY 241 VLVPDPTVSCLDPPHLPDQEDTPILEDSDSLEPNLSLAPBPVSGGLYIGIDDTELMDGAE 300
Db 241 VLVPDPTVSCLDPPHLPDQEDTPILEDSDSLEPNLSLAPBPVSGGLYIGIDDTELMDGAE 300
QY 301 KLPLXDSPIVSLDCPSLNNAFSLIADDSDTSTSIASPTSPVPLGESVLQDNSFDLN 360
Db 301 KLPLXDSPIVSLDCPSLNNAFSLIADDSDTSTSIASPTSPVPLGESVLQDNSFDLN 360
QY 361 NGSDAEQEMETQSSDFPPLTQAPDQSSSTTLQHPATSPAVSPPTTSPAVSLVSPAAASP 420
Db 361 NGSDAEQEMETQSSDFPPLTQAPDQSSSTTLQHPATSPAVSPPTTSPAVSLVSPAAASP 420
QY 421 ETSPEVCPAASTVSPVSPVSPASSAVLPAVSLVPLTASVTSFKASPVTSPPAAAPPT 480

Db 421 ETSPEVCPAASTVSPVSPVSPASSAVLPAVSLVPLTASVTSFKASPVTSPPAAAPPT 480
QY 481 ASPANKDVSSFLETTADVEEITGEGLTASGSDGVMRRRIATPEEVRLPLQHGWRREVRIK 540
Db 481 ASPANKDVSSFLETTADVEEITGEGLTASGSDGVMRRRIATPEEVRLPLQHGWRREVRIK 540
QY 541 KGSHRWQGETWYGP CGRMMKQFPEVIKYL SRNLVHSVRREHFSFSPRMPVGDPEERDT 600
Db 541 KGSHRWQGETWYGP CGRMMKQFPEVIKYL SRNLVHSVRREHFSFSPRMPVGDPEERDT 600
QY 601 PEGLOWVOLSABEETPSRIQAITGKGRPRNTEKATKEVPKVRGRGPPPKVKITELNKK 660
Db 601 PEGLOWVOLSABEETPSRIQAITGKGRPRNTEKATKEVPKVRGRGPPPKVKITELNKK 660
QY 661 TDNRLPKKLEAQETLINEEDKAKIAKSKKMRQKVQORGECLTTIQOARNKRKQETKSLKH 720
Db 661 TDNRLPKKLEAQETLINEEDKAKIAKSKKMRQKVQORGECLTTIQOARNKRKQETKSLKH 720
QY 721 KEAKKSKAEKEGKTKQEKLEKVKREKSKVK-KEKEEVTKAKPACKADKTLATQRRLL 780
Db 721 KEAKKSKAEKEGKTKQEKLEKVKREKSKVK-KEKEEVTKAKPACKADKTLATQRRLL 779
QY 781 EERQKQOMILEMCKPTEDMCLTDHOPLPDPSRVPGTLPSGAFSDCLTIVEFLHSFGKV 840
Db 780 EERQKQOMILEMCKPTEDMCLTDHOPLPDPSRVPGTLPSGAFSDCLTIVEFLHSFGKV 839
QY 841 LGFDPKADVPVSLGVLQEGLLCQGDSLGEVQDILLVRLKKAALHDPGPPSYCQSLKILGEKV 900
Db 840 LGFDPKADVPVSLGVLQEGLLCQGDSLGEVQDILLVRLKKAALHDPGPPSYCQSLKILGEKV 899
QY 901 SEIPLTRDNVSEIILRCFLMAYGVAPALCDRLRTPQFAQPPQQAALVAFPVHEUNGSTL 960
Db 900 SEIPLTRDNVSEIILRCFLMAYGVAPALCDRLRTPQFAQPPQQAALVAFPVHEUNGSTL 959
QY 961 IINEIDKTLSEMSVSYRKNKWIVEGRRLKTVLAKTGRSEVEMGRPEECGLRRSSRTM 1020
Db 960 IINEIDKTLSEMSVSYRKNKWIVEGRRLKTVLAKTGRSEVEMGRPEECGLRRSSRTM 1019
QY 1021 EETSGMEEEEEESIAAVPGRRRRDGEVDATASSIPELEKQIEKLSKQFRRKJLHLS 1080
Db 1020 EETSGMEEEEEESIAAVPGRRRRDGEVDATASSIPELEKQIEKLSKQFRRKJLHLS 1079
QY 1081 SOMLRVSLGQDQYRRRYWVLPYLAGIFVEGTEGNLVPEEVIKKTDSLKVAHAASLNP 1140
Db 1080 SOMLRVSLGQDQYRRRYWVLPYLAGIFVEGTEGNLVPEEVIKKTDSLKVAHAASLNP 1139
QY 1141 LFSKMEIAGNTTASSPARASRPLKTKPGFMQPRHFKSPVRGODSEOPQAOLOPEAQL 1200
Db 1140 LFSKMEIAGNTTASSPARASRPLKTKPGFMQPRHFKSPVRGODSEOPQAOLOPEAQL 1199
QY 1201 HVPAPQPOLQLOQSHKGFLQEGLSGLSOSQHDLSQSASFSLWSLSTQSHSSLLSSSV 1260
Db 1200 HVPAPQPOLQLOQSHKGFLQEGLSGLSOSQHDLSQSASFSLWSLSTQSHSSLLSSSV 1258
QY 1261 LTPDSSPKGLDPAPSQPPEPEPEDESSPDLQAFWFNISAQMPNAAATPPPLAVSEDP 1320
Db 1259 LTPDSSPKGLDPAPSQPPEPEPEDESSPDLQAFWFNISAQMPNAAATPPPLAVSEDP 1318
QY 1321 TFSPOQLASSKPMNRPSAANPCSPVQFSSTPLAGLAPKRAGDPCGEMPSPTGLCQPKR 1380
Db 1319 TFSPOQLASSKPMNRPSAANPCSPVQFSSTPLAGLAPKRAGDPCGEMPSPTGLCQPKR 1378
QY 1381 GRPSKFFKQEWQRYLTOLTAQPPVPEMCSGWWIPDPEMLDAMLKALHPGRIREKALHK 1440
Db 1379 GRPSKFFKQEWQRYLTOLTAQPPVPEMCSGWWIPDPEMLDAMLKALHPGRIREKALHK 1438
QY 1441 HLNKHRDFLOEVCLRPSADPIPEPRQLPAFOEGINWSPKETKTETDLAVLQWVELEQR 1500
Db 1439 HLNKHRDFLOEVCLRPSADPIPEPRQLPAFOEGINWSPKETKTETDLAVLQWVELEQR 1498
QY 1501 VIMSDLOIRGWTCTSPDSTREDLAYCEHLSQSDIETWRGPGREGLAQKRTTNPLDLAV 1560
Db 1499 VIMSDLOIRGWTCTSPDSTREDLAYCEHLSQSDIETWRGPGREGLAQKRTTNPLDLAV 1558

```
QY 1561 MRLAALQNVKRRYLREPLWPTHEWVLEKALLSTENGAGEGTTTISYEITPRIRIMROT 1620
Db 1559 MRLAALQNVKRRYLREPLWPTHEWVLEKALLSTENGAGEGTTTISYEITPRIRIMROT 1618
QY 1621 LQRCRAAHVCLCLGHLERSIAWEKSVNKNVTCLVCRKGDNDDEFLLCDGCDRGCHYCHR 1680
Db 1619 LQRCRAAHVCLCLGHLERSIAWEKSVNKNVTCLVCRKGDNDDEFLLCDGCDRGCHYCHR 1678
QY 1681 PKMEAVPGDMFCTVCLAQVEGETQKPGPKGQKRGKSGYSLNFSFGDGRRRVLLKG 1740
Db 1679 PKMEAVPGDMFCTVCLAQVEGETQKPGPKGQKRGKSGYSLNFSFGDGRRRVLLKG 1738
QY 1741 RESPAAGPRYSEERLSPSKRRRLSRNHHSDLTFCETILMESHDAAMPKXPENRPLV 1800
Db 1739 RESPAAGPRYSEERLSPSKRRRLSRNHHSDLTFCETILMESHDAAMPKXPENRPLV 1798
QY 1801 SGYRRIIKNPMDFSTMRELLRGGYTSSEEFADALLVFDNCQTFNEDDSEVGKAGHIMR 1860
Db 1799 SGYRRIIKNPMDFSTMRELLRGGYTSSEEFADALLVFDNCQTFNEDDSEVGKAGHIMR 1858
QY 1861 RFFESRWEFFYQKQANL 1878
Db 1859 RFFESRWEFFYQKQANL 1876

RESULT 5
US-09-418-710-21
; Sequence 21, Application US/09418710
; Patent No. 6596482
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042001
; CURRENT APPLICATION NUMBER: US/09/418,710
; CURRENT FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 21
; LENGTH: 1972
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-418-710-21

Query Match 25.3%; Score 2482.5; DB 4; Length 1972;
Best Local Similarity 32.2%; Pred. No. 1.8e-185;
Matches 680; Conservative 329; Mismatches 678; Indels 426; Gaps 68;

QY 1 MEMANEANDHFNFTGLPPAPAAAGLKPSPS-----SGEGLYTNKSGPMNFPQ 48
Db 43 MESSNSUSD-----SGTSSDTSSEGISDSDDDLEEDDEEDQDSEEDDDSDSESAQ 98
QY 49 GKSLNGDVNAGLSTVSHTTTSGILNSAPHSSTSHLHP--SVAYDCLMNYSQYPSANPG 107
Db 99 HKS--NNQVLLHG---ISDPKADG--QKATEKAQEKRIHQPPLAPE-----SQTHSFQ-- 145
QY 108 SNLKDPPLLSQ-----FGGQYPLNGI-----LGSQRQPS----- 138
Db 146 SQOKQPVLSQOLPPFIPOSSQAKESVNKHTSVIQTGLSVNVRPLSLVNOAKKETYMKL 205
QY 139 --PSHTNLRAGSOKFWANGTHSPMGLNFDSDQELYDSPDO-----NFEVCVSGI-- 186
Db 206 IVPSPDV--LKAGNKMTSESSLLTSELRSKEQYKQAFPSQLKQESKSLKKVIALSN 264
QY 187 -----HPDEAEKEMTSSVAENGTLGVCSLELEEXQPELKNCMGYNGSVPSLSHQ 237
Db 265 PKATSSSPAHPKQTLNNHPNPLTNAL-----LGNHPQ-----NG---VIQSVIQ 307
```

```
QY 238 EVSVLVPDPTVSCLDLDPHLPDQLEDTPILSEDSLEPFNSLAPBPVSGGLYGIIDDTELM- 296
Db 308 EAPL-----AUTTKMQSKINEN--TAAASSTFPSPVNLSTSGRRTPGQTPVMP 357
QY 297 -----GAEDKLPLXGDSPIVIALDCPSLNNAFSLALLADDSQTSITSIFASPTSPVL 347
Db 358 SASPILHSQKKEKAVSNVNPVKI-----QHHSHPAKSLVEQPRGTDSDIFSSKXSDSN 412
QY 348 GESVLQNSFDLNNGSAAQOEEMETQSDPFPPLTQPA-----PDQSTTQLHPAT 398
Db 413 EDEBEDDEEDDEDEDDSDSQSDSDNSSESDETEGSEEDDDDDKQDQSDSDTEGEXT 472
QY 399 SPAVSPITSPAVSLVSPAASPEISPEVCPAASTVSPVSPVSPVSPVSPVSPVSPVSPVSPV 458
Db 473 SMKUNKTIS-----SVKSPSMS----- 489
QY 459 LTASVTSPKASPVTPSPAAAFPTASPANKDVSSFFLETTADVVEITGEGLTAS--GSGDVMR 517
Db 490 LTGHSTPRNLHIAKAPGSA--PAALCSSESQSPAFIGTSSST-----LTSSPHSGTSKRR 541
QY 518 RIATPEVRLPLQHGWRREVRIRKKGSHRWQGETWYQPCGKRMKQFPEVIKYLRLNVHS 577
Db 542 RVTDERELRIPLEYQWQRETRIRNFGGRLOQEVAYAPCGKKLQRYPEVIKYLGRNGIMD 601
QY 578 VREHFSPSPMPVGDFFFEERDTPGLOWQLSAEEIPSRIOAITGKRGPRNTEKATK 637
Db 602 ISDNFSAKIRVGDIFYEARDGQPMQWCLKEEDVIPRAMEGRGRPNPDQRAR 661
QY 638 EVPKVKGRRGRRPPVKITELINKTNDNPLKKLEAQE-----TLNEEDKAKIAK 685
Db 662 EESMRRRKGRPPNVGNAEFLDNADAKLLRLKLOAQEITARQAAQIKLLRLKLOQBOARVAK 721
QY 686 SKKMR-----QKVRGEGCLTTIQOQARNKRKQETSLEK-----KEAKKXSAEK 733
Db 722 EAKKQQAIAAAEKKQKEQIKMKQOEKIKRIOIRMEKELRAQQLLEAKKKEEAAAN 781
QY 734 GK--TKQKLEK-----VKREKKEKVM-----KEKEEVTAKPACAD 771
Db 782 AKLLEAKRIKEKEMRQQAALLKHQERRRRQHMLMKAMEARKKAAEERLQKERRDE 841
QY 772 KTLATORLEERQKQOMILEMKKPTEDMCITDHPPLDFPSRVFPLTPSGAFSDCLTIV 831
Db 842 KLLNKERKLEQRLELEWAKELKKNEDMCLADQKPLPELPRIFGLVLSGTSFSDCLMVV 901
QY 832 EPLHSFGVLGFDPAKDVPSGLVLOEGLCOGDSLGEVQDILLVRLKLAALHDPGFPSCYQ 891
Db 902 QFLNFGKVLGFDVNIIDVPNLVLOEGLLNTIGDSMGEVQDILLVRLLSAAVCDPGLITGYK 961
QY 892 SLKTLGKGVSEIPLTRDNVSEILRCFILMAYGVXPALCDRLRTOPFOAOPPOKAAVLAFP 951
Db 962 AKTALGHLNVGNVRNVNSEILQIFMEAHCGQTELTESLTKAFQAHPTPAKASVLAPL 1021
QY 952 VHELNGSTLIINEIDKLTESMSSYRKNKIWVEGRRLKTLVAKRTGRSE-----VEMGRP 1007
Db 1022 INELACSKSVSEIDKNIDYMSNLRRDKWVVEGKLRLRIHAKTKRDTSGGIDLEE 1081
QY 1008 BECLG-----RRSRRIWE-----ETSGNEEEEEEESTAAVPRGRGR--RDEGV 1049
Db 1082 QHPLGTPTPKRRRKRGKGDSDYDDDDDDDDQDQDEDEDEDEDEKDGKKTDCEDDE 1141
QY 1050 DATASSIPELEQLEKLSKQLPFRKLLHSSQMLRAVSLGQDRYRRYVLPVLAGIFV 1109
Db 1142 GDQAASVEELEKQLEKLSKQSQYRRKLFDAHSLSRVNMGPDYRRYVLPVLCGGGLFV 1201
QY 1110 EGTENGLVPBEVIKKTDSLKVAHAASLNPALEFMSKMLAGSNNTASSPARARS--PLKT 1168
Db 1202 EGMESGEGLEE--IAKEREKUKAESVOIKEEMF-----ETSGDSLNCSTNDTCEKEDLKE 1256
QY 1169 KPG---FMQ--PRHPKSPVRGQDSEQQAQLOPEAQLHVPQAQPOQLQLQLOSHKGFLEQ 1224
Db 1257 KDNTNLFLQKPGSPSKSLKLE-----VAKMPPESEVMTP-----KPNAGAN 1298
```

Qy	1225	GSPLSLGQS-QHDLS--QSAFLSWLSQTSQSHSLSSSVL--TPDSSFGKL-DPAPEQPP	1278
Db	1299	GCTLSYQNSGKSLGSVQS-----TATQSNVEKADSNLFTNGSSGFGKFSPLPND--	1350
Qy	1279	EEPEPDEAESPDLOAQAFWNISAOBPCN-----AAPTPEPLAVSEDOPTSPQQ	1326
Db	1351	-----QLLKTLTEKRWQFSLPRTPCDDTSLTHADMTASLVTTPSQPSKSFSPPTAP	1405
Qy	1327	LASGKPNRPSAAMP--CSPVQF--SSTPLAGLAPKRRAGDPG-----EMPOSPT	1372
Db	1406	LGSS--AQNPVGLNPFALSPLOVKGVSMMGL---QFCGWPTGVVTNSIPTLTSVPSLGS	1460
Qy	1373	GLGQPKRGR-----PPSKF---FKMQQRVLTQTLTA-----QPVPPEMCS	1410
Db	1461	GLGLSEGNNGSFLTSNVASSKSESPVQNEKATSAQPAAVEVAKPVPFSPSPKPIPEMQF	1520
Qy	1411	GWMIIPDEMLDAMLKALHPRGIREKALHKHLNKHDRDFLOEVCULRPSADPIFEPRQLPAF	1470
Db	1521	GWRIIDPEDLKALKVLHLRGIREKALQKQIQKHLDYITQACLKXKDVAIIELENEEN	1580
Qy	1471	Q---EGIMSWSPKETYETDLAVLQWVEELEQRVIMSDLOIRGWTCPSPDSTREDLAYCE	1527
Db	1581	QVTRDIVENNSVEQAMEMDLSVLQVEDLERRVASASLOVKGWMCPEPASEREDLVYFE	1640
Qy	1528	HLSDS-----QEDITWRQFGRGL-APORKTNPULDLAWMRLALAEQNVKRYRLREPLUMPT	1582
Db	1641	HKSFTKLCKEHDGEFTGEDESSAHALERKSDNPLDIATVRLADLERNIERR-----	1691
Qy	1583	HEVVLERKALLSTPNGAPEGITTEISYEITPRIRLWRQTLQRCBSAAHVCLCLGHLERSIA	1642
Db	1692	-----FEEDIAPGLRVRRRLSEARSAAQVALCICQOLQKSIA	1728
Qy	1643	WEKSVNKTCLVCRKGDNDLFLLICDGCGRGCHYIHRPKMEAVPEGDWFTCVCLAQ---	1699
Db	1729	WEKSIMVYCOICRKGDNEELLILCDGCDKGCHYIHRPKITIPDGDWFCPACIAKASG	1788
Qy	1700	-----QVEGETPQKGPGRGQK-----RKSGVSYLNFSEGDGRRRRVLLK	1739
Db	1789	QTLKIKLHLVKGKKTNB---SKGKKVTLTGDTEDEDSASTSSSLKRGNKDLQKRM---	1842
Qy	1740	GRESPAAGPRYSERLSPSKRRRLSMENHSDITFCFIILMEMESHDAAMPPEXPVNPRL	1799
Db	1843	-ENTSINLSQBSFTSVKPK-----RDDS KDIALCMLITEMETHEDEAMPFLLPVNLKL	1897
Qy	1800	VSGYRRIIKNPMDFSTWRERLLRGYTSSEBEAADALLVFDNQTFNEDDSEYVKAGHIM	1859
Db	1898	VPGYKKVIXKPMDFSTIREKLSGQYENLETFALDVLVFDNCTFNEDDSDIGRAGHNM	1957
Qy	1860	RRPFESRWEBFYQ	1872
Db	1958	RKYFEKKWTDTFK	1970

RESULT 6

```

US-09-839-479-21
; Sequence 21, Application US/09839479
; Patent No. 672722
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042002
; CURRENT APPLICATION NUMBER: US/09/839,479
; CURRENT FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 72
SOFTWARE: Fast-Seq for Windows Version 4.0

```



```
; CURRENT APPLICATION NUMBER: US/09/690,454
; CURRENT FILING DATE: 2000-10-18
; PRIOR APPLICATION NUMBER: 09/189,144
; PRIOR FILING DATE: 1998-11-10
; PRIOR APPLICATION NUMBER: 60/044,039
; PRIOR FILING DATE: May 30, 1997
; PRIOR APPLICATION NUMBER: 60/048,093
; PRIOR FILING DATE: May 30, 1997
; PRIOR APPLICATION NUMBER: 60/048,190
; PRIOR FILING DATE: May 30, 1997
; PRIOR APPLICATION NUMBER: 60/050,935
; PRIOR FILING DATE: May 30, 1997
; PRIOR APPLICATION NUMBER: 60/048,101
; PRIOR FILING DATE: May 30, 1997
; PRIOR APPLICATION NUMBER: 60/048,356
; PRIOR FILING DATE: May 30, 1997
; PRIOR APPLICATION NUMBER: 60/056,250
; PRIOR FILING DATE: August 29, 1997
; PRIOR APPLICATION NUMBER: 60/056,296
; PRIOR FILING DATE: August 29, 1997
; PRIOR APPLICATION NUMBER: 60/056,293
; PRIOR FILING DATE: August 29, 1997
; NUMBER OF SEQ ID NOS: 229
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 94
; LENGTH: 196
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-690-454-94

Query Match      10.5%; Score 1030; DB 4; Length 196;
Best Local Similarity 98.5%; Pred. No. 7.9e-73;
Matches 193; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1683 MEAVEGDMFCTVCLAQVGEFTQKPGPKRGQKSGYSLNFSEGDGRRRLVLLKGR 1742
Db 1 MEAVEGDMFCTVCLAQVGEFTQKPGPKRGQKSGYSLNFSEGDGRRRLVLLKGR 60

Qy 1743 SPAAGPRYSERLSPKRRRLSMRHHSDLTCEIILMEMESHDAAPFPXPVNPRLVSG 1802
Db 61 SPAAGPRYSERLSPKRRRLSMRHHSDLTCEIILMEMESHDAAPFPXPVNPRLVSG 120

Qy 1803 YRRIKPNMDFSTMRERLLRGYTSSEFAADALLVFNQCFTFNEDDSEVKGAGHIMRRF 1862
Db 121 YRRIKPNMDFSTMRERLLRGYTSSEFAADALLVFNQCFTFNEDDSEVKGAGHIMRRF 180

Qy 1863 FESRWEFFYQKQANL 1878
Db 181 FESRWEFFYQKQANL 196

RESULT 10
US-09-418-710-65
; Sequence 65, Application US/09418710
; Patent No. 6596482
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042001
; CURRENT APPLICATION NUMBER: US/09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 65
; LENGTH: 175
; TYPE: PRT
; ORGANISM: Homo sapiens

Query Match      8.9%; Score 875.5; DB 4; Length 175;
Best Local Similarity 97.7%; Pred. No. 9.4e-61;
Matches 172; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

Qy 780 LEERQKQOMILEEMKKPTEDMCLTDHQPLPDFSRVPGTLTLPAGFSDCLTIVEFLHSFGK 839
Db 1 LEERQKQOMILEEMKKPTEDMCLTDHQPLPDFSRVPGTLTLPAGFSDCLTIVEFLHSFGK 60

Qy 840 VLGFDPADKDVPSLGVQLQEGLLCQGSLSGEVODLLVRLKKAALHDPGFFSYCOSLKILGEK 899
Db 61 VLGFDPADKDVPSLGVQLQEGLLCQGSLSGEVODLLVRLKKAALHDPGFFSYCOSLKILGEK 120

Qy 900 VSEIPLTRDNVSEILRCFLMAYGVKXPALCDRLRTQPPQAAVLAPFVHEL 955
Db 121 VSEIPLTRDNVSEILRCFLMAYGVKXPALCDRLRTQPPQAAVLAPFVHEL 175

RESULT 11
US-09-839-479-64
; Sequence 64, Application US/09839479
; Patent No. 6727222
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042002
; CURRENT APPLICATION NUMBER: US/09/839,479
; CURRENT FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 64
; LENGTH: 175
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-839-479-64

Query Match      8.9%; Score 875.5; DB 4; Length 175;
Best Local Similarity 97.7%; Pred. No. 9.4e-61;
Matches 172; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

Qy 780 LEERQKQOMILEEMKKPTEDMCLTDHQPLPDFSRVPGTLTLPAGFSDCLTIVEFLHSFGK 839
Db 1 LEERQKQOMILEEMKKPTEDMCLTDHQPLPDFSRVPGTLTLPAGFSDCLTIVEFLHSFGK 60

Qy 840 VLGFDPADKDVPSLGVQLQEGLLCQGSLSGEVODLLVRLKKAALHDPGFFSYCOSLKILGEK 899
Db 61 VLGFDPADKDVPSLGVQLQEGLLCQGSLSGEVODLLVRLKKAALHDPGFFSYCOSLKILGEK 120

Qy 900 VSEIPLTRDNVSEILRCFLMAYGVKXPALCDRLRTQPPQAAVLAPFVHEL 955
Db 121 VSEIPLTRDNVSEILRCFLMAYGVKXPALCDRLRTQPPQAAVLAPFVHEL 175

RESULT 12
US-09-418-710-1
; Sequence 1, Application US/09418710
; Patent No. 6596482
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042001
; CURRENT APPLICATION NUMBER: US/09/418,710
; CURRENT FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
```

```

; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1674
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-418-710-1

Query Match      6.8%; Score 671.5; DB 4; Length 1674;
Best Local Similarity 19.8%; Pred No. 4.4e-43;
Matches 305; Conservative 197; Mismatches 399; Indels 639; Gaps 44;

Qy 644 RGRGPPKVKITELNKTNDNPLKLEAQETLNEEDKAKIAKSKKQKQVQGECLTTI 703
Db ||||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Qy 704 QGOARNKQKQETSLKHAEKKSXAEBKGTQKQKLEKVKREKKEKVKQKKEBVTK 763
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 471 AFEKAKLRKADAL---EAKKKEKEDKEX---KREELKKIYEEB---RLKKKEE--- 516
Qy 764 AKPACKADKTLTATORLEEROKQMILEMKKPTEDMCLTDHOPLPDPSRVPGLTLPSCA 823
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 517 -KERLKVREKREKLEKREKRYEYLKQWKPREDMCEDDLKELPEPTPVK--TRLPPEI 574
Qy 824 FSDCLTIVFHLHSGKVLGFPDPAKVP---SLGVLEGLLQGGSLGSEVODLLVRLKAA 880
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 575 FGDALMWLEFLNAGEL--FDLQDFPDGVTLVLEALV--GNDSEGLPCELLFFFLTAI 631
Qy 881 LHDGPPSYCOSLKLGLKGVSEIPLTRDN-----VSEILRCLFMAYGV-- 923
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 632 F-----QAIAEEEEVAKEQLTDADTKGCSLKSLDLSCTLSEILRLHLILASGADV 682
Qy 924 -----XPALCDRLTQPFQAQPPQOKAAVLAFFVHEL 955
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 683 TSANAKYRYQKRGFPDATTACMEURLNPNLNVKLSSTSVYDLTPGEMKIL-----NAL 738
Qy 956 NGSTL-----IINEIDKLTLESMS--YRKNKVIWEGRRLRLKTVLAK----- 995
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 739 CGKLLTLVTRDFIEDYVILRQAQKQFRELKAEQHRKEREAAAIRKRKEEKLKEQEQ 798
Qy 996 -----RTGRSEVMGRPE-----ECLGRRSSRIMEETSGMBEEEEESTAA 1037
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 799 KXKKEQKLEKEDQENSTADISIGBEEREDFTSIESKDTEQKELDQDMFTEDDEDPGSH 858
Qy 1038 VPGRRGRDGEVDATASSIPELERO-----IEKLSKRLQFPFRKLLHSSQ 1082
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 859 XGRGRKKG-----QNGFKEFTROEQINCVTRRELLTADBEALKEQHRKEKELLEKIQ 912
Qy 1083 MLRAVS-----LGQDRYRRYVLPYLAGIFVEG-----TEGNLPEEVIKKETDSLKVA 1133
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 913 SAIACNTNIFPLGRDRMYRYMIFPSIPGLFIEDYSGLTEDMLLP----- 957
Qy 1134 HASLNPAULFSMKMELAGSNTTASSPARARSPLTKPGFMOPRHPKSPVRGQDSEQPOAQ 1193
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 958 -----RPSSFQNNVQSQD----- 970
Qy 1194 LQPEAQLHVPAPQPOLQIQSHKGFLEQBSPLSLGCSQHDLSQSFLSLWSLQTSQS 1253
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 971 -----PQV-----STKGE 979
Qy 1254 SLLSSVLTPSPGKLDPAKSPPEPEDEAESSPDQAFWFNISAQMPCNAAPTPL 1313
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 980 PLMSESTNIDQGR-----DHSVQLP----- 1001
Qy 1314 AVSEDOPTSPQQLASSPKMNRPSAANPCSPVQSSSTPLAGLAPKRRAGDGPMPQSTG 1373
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1002 -----KPVHKPNR----- 1009
```

```

Qy 1374 LGQPKRRGRPPSKFFKQMEQRYLTQLTAQVPPPEMCSGWWIIPDEMLDAMLKALHPRGI 1433
Db -----WCYSSCEQLDQIEALNSRGL 1031
Qy 1434 REKALHKLHNRDFOEVCLRPSADPIF-----EPRQLPAPQEGIMS--WSPKEXTVE- 1485
Db ||||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Qy 1032 RESALKETLLQEKSRICAQLARFSEKFFHSDKQPDSPKPTYSRGRSSNAYDPSQCAEK 1091
Qy 1486 -TDLAVLQWVEELEQRVIM-----SDLQIRGWTCPSPD-----STRE 1521
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1092 QLELRLDLFDLIEDRIYQGTGLGAVTDRIHWSALESGRYELLSEENKENGIIKTYNE 1151
Qy 1522 DLAYCEHLSDSQEDITWRGPGREGIAPQKTTNP-----LDLAVMLAALEQNVKR 1572
Db : : : : : : : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Qy 1152 DVEEMIDEQTKVVKDRLGIKTKETPTSTVSTNASTPOSVSSVHYLAKALFQIEQGIER 1211
Qy 1573 RYLREPL-----WPTHEVLEKALLSTPNGAPEGTTTETSYEITPRIRWQTLQRCRSA 1627
Db : : : : : : : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Qy 1212 RFLKAPLDASDGRSYKTVLDR-----WRESLLSSASL 1244
Qy 1628 AHVCLCLGHLERSIAWKSNNKVTCLVCRKGDNDDEFLLLCDGDRGCHIYCHRPQMEAVP 1687
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1245 SQVFLHLSTLDRSVIWSKILNARCKICRKGDAENMVLCGDCDRGHHTYCVRPKLKTPV 1304
Qy 1688 EGDWFTVCLAAQ-----VEGEFTQKPGFKRGQRKSGYSLN 1725
Db ||||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Qy 1305 EGDWFCPECRPKQRCRLSPRQPSLESDVEDSMGDEDDVDGDEEGQSEEEYEVE 1364
Qy 1726 FSEG-----GR-----RRRVLLKGR-----ESPAGAPRIYE----- 1752
Db : : : : : : : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Qy 1365 QDEDDSOEEERVSPLKRGQPVRLPVKTRKGLSSFSRSGQOQBPFGRYPSRSQOSTPKTT 1424
Qy 1753 -----ERLSPSKERR----- 1762
Qy 1425 VSSKTGRSLRKINSAPPTETKSLIASRSTRSHGHPLOADVVELLSPRRKRRGRKSANN 1484
Qy 1763 -----LSMNHHS----- 1770
Qy 1485 TPENSPNFPNFRVIATKSSQSRSVNIATSKLSLOESESKRCKRQSPSPVTLGRSS 1544
Qy 1771 -----DLTCEIILMEMESHDAWPFKEPVNPRLVSGYRRIIKNPMDFSTMERLLRG 1823
Db : : : : : : : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Qy 1545 GRQGVHELSAFEQVAVELVRHDDSWPFLKLVSKIQVDPDYDIKKPIALNIIREKNKC 1604
Qy 1824 GYTSSEFAADALLVFQNCQTFNEDDSEVGKAGHIMRFF 1863
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 1605 BYKLASEFIDIELMFNCFEYNPRINTSEAKAGTRLQAF 1644

RESULT 13
US-09-839-479-1
; Sequence 1, Application US/09839479
; Patent No. 6727222
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042002
; CURRENT APPLICATION NUMBER: US/09/839,479
; CURRENT FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1674
; TYPE: PRT
; ORGANISM: Homo sapiens
```



```
Qy 764 AKPACKADKTLATORRLEEROKQOQMILEEMKKPTEDMCLTDHQPDPFSRVPGTLPSGA 823
Db 525 EK-----EREKLEERKKYVYKQWKPREDMECDLDELPEPTPVK--TRLPEI 573
Qy 824 FSDCLTIVEFHSFGKVLGFPDPAKVP---SLGVLQEGLLCGDGLGEVODLLVRLKAA 880
Db 574 FGDAIMVLEFLNAGEL--FDLODEFDPGVTLVEEALV--GNDSEGFLCELLFFFLTAI 630
Qy 881 LHDPGPPSYCOSLILGEKVSEIPLTRDN-----VSEILRCFLMAYGV-- 923
Db 631 F-----QAIABEEEEVAKBQTLADTKGCSLSLSDLSCTLSEILRLHILASGADV 681
Qy 924-----XPALCDRLTOPFOAQPPQQAALVLAFFVHEL 955
Db 682 TSANAKYRYQKRGGFATDDACMEURLSNPLVKLSSTSVYDITPGEMKIL-----HAL 737
Qy 956 NGSTL-----IINEIDKTLESMS--YRKNKWIVBGRRLRLKTVLAK----- 995
Db 738 CGKLLTLVSTRDFIEDYVILRQAKQEPRELKABQHRKEREAAAIRKRKEEKLKQEQ 797
Qy 996-----RTGRSEVEMGRPE---ECLGRRRSRIMEETSQWEEEBEESIAA 1037
Db 798 KMKKEQKELKEDEQRNSTADISIGEEREEDFTSIESKDTQKELDQDMFTEDDDPGSH 857
Qy 1038 VPRGRGRDGEVDATASSIPELQ-----TEKLSKQLFFERKLLHSSQ 1082
Db 858 KRGARGKG-----QNGKFEFTQEQINCVTRELLTADEEALKQEHQRKEKELLEKLQ 911
Qy 1083 MLRAVS-----LGODRYRRRYVWLYLAGIFVEG-----TEGNLVPEEVIKKTDSLKVAA 1133
Db 912 SAIACNIFPLGRDRMYRYWIFSLPGLFIEEDYSGLTEDHLLP----- 956
Qy 1134 HASLNPAFLFSWKMELAGSNTTASSPARARSPLTKPGFMOPRHFSPVRQDSEQPAQ 1193
Db 957-----RPSFQNNVQSQD----- 969
Qy 1194 LQPEAQLHVPAPQPOLQLOQSHKGFLEQBSPLSLGQSQHDLSQSAFLSMLQTSOHS 1253
Db 970-----PQV----- 978
Qy 1254 SLLSSSVLTPDSSPKLDPAPSPQPEPEPDEABSSPDQAFWFNISQAQPCNAAPTPL 1313
Db 979 PLMSESTSNIDQGR-----DHSVQLP----- 1000
Qy 1314 AVSEDOPTSPQQLASSKPMRNPAAANPCSPVQFSSTPLAGLAPKRRAGDPGEMPQPTG 1373
Db 1001-----KPVHKPNR----- 1008
Qy 1374 LGQPKRGRPPSKFFKQMEQRYLTQLTAPVPPMCSGWWIPDPPEMLDAMLKALHPGI 1433
Db 1009-----WCFYSCEQLDQIEALNSRGH 1030
Qy 1434 REKALHKLHNRDFLOEVLCLPSADPIF-----EPRQLPAFQEGIMS--WSPKRETYE- 1485
Db 1031 RESALKETLLQESRICAQLARFSEEFHFSDEKQDPDSKPYISGRSSNADPDSQMAEK 1090
Qy 1486 -TDLAVLQWBEELQORVIM-----SDIQIRGWTCPSPD-----STRE 1521
Db 1091 QLELRDLFDLIDRIYQGTGLGAVKVTDRHIWRSALSGRYELLSEENKENGIIKTVNE 1150
Qy 1522 DLAYCEHLSDSQEDITWRGPGREGLAPOKTTNP-----LDLAVMLAALAEONVKR 1572
Db 1151 DVEEMEDQTKVVKORLLGIKTETFTSTVNASTPQSIVSVVHYIAMAUFQIEQGLER 1210
Qy 1573 RYLREPL-----WPTHEVLEKALLSTPNGAPEGTTTIEISYEITPRIRWQTQRCRSA 1627
Db 1211 RFLKAPLDASDSRSYKTVLDR-----WRESLLSSASL 1243
Qy 1628 AHVCLCGLHLERSIAWEKSVNKNVTCVCRKQDNDEFLLLCGDGRGCHYIHRPKRBAVP 1687
Db 1244 SQVFLHLSTLDRSVIWSKSIILNARCKICRKGDAENMVLCDCDGRGHHTYCVRPKLKIVP 1303
```

```
Qy 1688 EGDWFCVCLAAQ-----VEGEFTQKPGFKRQKQKSGYSLN 1725
Db 1304 EGDWFCPECPKQRCRLRSFRQBSLSDDEVEDSDMGDEDDVDGDEEGQSEEEYEVE 1363
Qy 1726 FSEGQ-----GR---RRRVLLKGR-----ESPAAGPRYSE----- 1752
Db 1364 QDEDDSQEBEEVSLPKRGRPOVRLPVKTRGKLSRSSFSRGQQQFGRYPYRSQOSTPKTT 1423
Qy 1753-----ERLSPSKERR----- 1762
Db 1424 VSSKTRSLRKINSAPPTETKSLRIASRSTRHSHGPIQADVFBVLLSPRRKRGRKKSANN 1483
Qy 1763-----LSMENHHS----- 1770
Db 1484 TPENSPNFPNPRVIATKSSSQSRSVNIASKLSLOESKRCRCRQSPSPVTLGRSS 1543
Qy 1771-----DLTCEILMEMESHDAAMPFEPVNPRLVSGYRRIIKNPMDPFTMRERLLRG 1823
Db 1544 GRQGVHELSAFEQLVVELVRHDDSWPFLKLSKIQVPDYDIKKPILALNIIREKVNKC 1603
Qy 1824 GYTSSEFAADALLVFDNCQTFNEDDSEVGKAGHIMRRFF 1863
Db 1604 EYKLASEFIDDIELMFNSCFEYNPRNTSEAKAGTRLQAFF 1643

RESULT 15
US-09-839-479-69
; Sequence 69, Application US/09839479
; Patent No. 6727222
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042002
; CURRENT APPLICATION NUMBER: US/09/839,479
; CURRENT FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 69
; LENGTH: 1673
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-839-479-69

Query Match 6.8%; Score 669; DB 4; Length 1673;
Best Local Similarity 19.7%; Pred. No. 7e-43;
Matches 303; Conservative 200; Mismatches 397; Indels 640; Gaps 43;

Qy 644 RGRPRPKVITELLNKTDRPLKLEAQETLNEEDAKIAKSKKKRQKQVQCECLTTI 703
Db 424 RRRGRPPK-RIH--ISQEDN-----VANKQTLASVRSKATKRDKLLKQEMKSL 470
Qy 704 QGOARNRKQETKSLKHKAACKSKSAEKGKTKQEKLBKVKREKKEKVKQKKEEVTK 763
Db 471 AFEKAKLREKADAL---EAKKEKEDKEK---KREELKTIVEERLKKKEERLKVVER 524
Qy 764 AKPACKADKTLATORRLEEROKQOQMILEEMKKPTEDMCLTDHQPDPFSRVPGTLPSGA 823
Db 525 EK-----BREKLREKRYKVEYKQWSKPREDMCEDDLKELPEPTPVK--TRLPEI 573
Qy 824 FSDCLTIVEFHSFGKVLGFPDPAKVP---SLGVLQEGLLCGDGLGEVODLLVRLKAA 880
Db 574 FGDAIMVLEFLNAGEL--FDLODEFDPGVTLVEEALV--GNDSEGFLCELLFFFLTAI 630
Qy 881 LHDPGPPSYCOSLILGEKVSEIPLTRDN-----VSEILRCFLMAYGV-- 923
```


RESULT 17
US-09-949-016-7037
; Sequence 7037, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7037
; LENGTH: 1540
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-7037

Query Match 5.4%; Score 528.5; DB 4; Length 1540;
Best Local Similarity 19.5%; Pred. No. 6.9e-32;
Matches 265; Conservative 179; Mismatches 439; Indels 479; Gaps 42;

Qy	619	QAITKGRPR--NTEKAKTEPKVKRGRGRPPKVKITEL-----LNKTDNRPLKL	669
Db	472	QKSTGNSKSPKGLKTPTKWKQMTLDMAGTQ---KXTRAPNSGGTPTTSKPKHKL	528
Qy	670	EA-----QETLNEEDKAKIAKSKKMKQKQVORGECLATTIOQARNEKQKQTSKLK	720
Db	529	PPAALHLIAYKENKREDKESALSVCISKARLLSSDDRRLPEELSLVQKRYELLEH	588
Qy	721	KE--AKKXSAEKEGKTKQSKLEKVKREKKEKVKQKKEVEVTKAPKADKTLATQR	778
Db	589	KRWASMEEQREKYLKKRBEKKLKEKAKER---REK-----	626
Qy	779	RLEERQKQMILEMKKPTDMCLTDHQPDSRVPGLT-LPSGAFSDCLTIVEFLHSF	837
Db	627	MLERLEKQ-----KRYEQELTG-KNLFAFLVDTPEGLPNTLPDGVAMVVEFLSCY	677
Qy	838	GKVLGFDPAKDVPSIGVLQGLLQCGSLGVBQDLLVRLKLAALHDPGFPSPYCQSLKILG	897
Db	678	SGLLLPDAQYPTAVS-LMEALSADKGGGLYLNRLVILLQTLQDEIAEDYGE---LG	732
Qy	898	EKVSEIPLTRONVSEILRCFLMAYG-----VXPALCDRLRTQ	934
Db	733	MKLSEIPLTHSVSELVRLCLRRSDVQSESGSDTDNDKDSAAAFEDNEVQDEFLEKLETS	792
Qy	935	PFQAPPOOKAVALFAPPHLNGSLTIINEIDKTLSSSYRKQKWIIVEGRLRLK----	990
Db	793	EFFELTSEEKQILTALCHRI-----LMTYSQDHMETRQQMSAELW--KERLAVLKEEND	846
Qy	991	TVLAKRTGRSEVEM-----GRPEECGLRRRSRMEETSGMBEESIAAIVPGRR---	1042
Db	847	KKBAEKQKKEWAKENGKGVENGLGKTDKKEIVKEPQVDVTAEDMIISAVKSRLLA	906
Qy	1043	---GRDGEVDVATASIPELERQIEKLSKR-----QLFRRKKLHSSQMLRAVSLGQDRY	1094
Db	907	IQAKEREIQEREMKV-KLERQABEERIRKHAARAKAFQEGIAKAKLVMRRTPIGTDRN	965
Qy	1095	RRRYWVL--PYLAGIFVECTEGLNVPEEVIKKETDLSKVAHAASLNPALFSMKMELAGSNT	1153
Db	966	HNRYWLFSDVEVPLFIE-----KGWVHDSID-----	991
Qy	1154	TASSPARARSRLTKPGFMQPRHFKS--PVRGQSEQFPQALQPEAQHLVPAQPOPQLQL	1212

Db	992	-----YRFNHCKDHTVSGDEDYCPRSK-----	1014
Qy	1213	QLQSHKGFLEQEGSPLSIGSQSOHDLSSQAFSLWSLQTSQSHSSLLSSSVLTDPDSPGKLDLP	1272
Db	1015	-----KANLGKN-----ASMTQHGTAATEVAVEITTPKQG-----	1044
Qy	1273	APSQPPEPEDEAESSPDLQAFWFNISAQMPCNAAPTPLPLAVSEDOQTPSPQQGLASSKP	1332
Db	1045	-----QNLWFLCDSQ-----	1054
Qy	1333	MNRPSAANPCSPVQFSTPLAGLAPKRRAGDPGEMPOSPTGLGQPKRRGRPPSKFFKOME	1392
Db	1055	-----	1054
Qy	1393	QRYLTQLTAQVPVPEMCSGWWIIPDPMLDAMLKALHPRGIREKALKHKLANK-HRDFLOE	1451
Db	1055	-----KELDELLNCLHPQIGRESQKLERLKRKYODIHS	1088
Qy	1452	VCLRPSADPIPEPQLPAFOEGIMSWSPKE-----KTYETDLAVLQWVELEQORVIMSD	1505
Db	1089	IHL--ARKPNLGKSKCDGNQE-LNLFRLSDLIIVATRLQKGLGVVEETSEFEARVIS--	1143
Qy	1506	LQIRGWCPSPDSTREDLAYCEHLSDSQEDITWRGPGREGLAPORKTTNPLDLAVMLAA	1565
Db	1144	-----LEKLKDFGECV-----IA	1156
Qy	1566	LEQNVKRYLREPLWP-----THEVLEKALLSTPNGAPEGTTTTSIYEIT	1611
Db	1157	LQASVIKFLQGFQWAPKQKRKLOSEDSAKTEEVEDEEKOMVE-----EAKVA	1203
Qy	1612	PRIRWQTLQRCRAAHVCLCLGHLERSIAWEKSVNKKVTCLVCRKGDNDSEFLLLCDGCD	1671
Db	1204	SALKWKTAIREAOTFRMHVLLGMLDACIKWDMSAENARCKVCRKKGDDKLLILCDECN	1263
Qy	1672	RGCHYCHRPQMEAVPEGDMFCTVC-----LAQVVEGEFTQKPGPK	1713
Db	1264	KAFHLFLRPALEYVDPGEMQCPACQATARNRGRNYTEESASEDESEDESEEEEE	1323
Qy	1714	RQQRKSGSYLNFSEGDRRRVLLKGRES---PAA--GPR-----YSEERLSPSK----	1759
Db	1324	EEEEEEEDYEV---AGLRLPRKTRIRGKHSVIPPAARSGRRPGKKPHSTRSQKAPPVD	1380
Qy	1760	-----RRRLSMRNHSDLTFCETILMEMESHDAAPFPXPNRPLVSGYRRIIKNP	1810
Db	1381	DAEVDLVLQTKRSRRRQSLQKCEILHKIVKYSWPFREPVTREADEADYDVITHP	1440
Qy	1811	MDFTMERLLRGYTSSEEFADALLVFDNCQTFNEDDSEV	1852
Db	1441	MDFTVQNKSCGYSYRQVEFLTDMKQVFTNAEVTNCRGSHV	1482

RESULT 18
US-09-418-710-69
; Sequence 69, Application US/09418710
; Patent No. 6596482
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042001
; CURRENT APPLICATION NUMBER: US/09/418,710
; CURRENT FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 69
; LENGTH: 1525
; TYPE: PRT
; ORGANISM: Homo sapiens

US-09-418-710-69

Query Match 5.3%; Score 521.5; DB 4; Length 1525;
Best Local Similarity 19.6%; Pred. No. 2.4e-31;
Matches 281; Conservative 185; Mismatches 444; Indels 523;

Qy	565	EVILYLSRNLVHSVRHPSFSPMP-----VGDFEERTDPBGLQWQVLSABEIPS	616
Db	368	EMKMOSPNKLT-----NPHIPKGGPPAKPG---KHSDKP-----LKAGRSK	409
Qy	617	RI---QAITKGRGRPR--NTERAKTKEVPKVRGRPRPKVITEL-----LNKTDNR	664
Db	410	GILNGKSTGNSSPKKGLKTPTKMKOMTLLDWAQGTQ---KXTRAPNSGGTPTSSK	466
Qy	665	PLKKLEA-----OETNEBDSKAKIAKSKKOROKVORGECLTITTOQOANRKRQET	715
Db	467	PHKHLPPAALHLIAYYKENDKREDKRSALSCVISTKARLLSSEDRARLPPELRSLVQRY	526
Qy	716	KSUKHKE--AKKKSXAEBKGKTKOEBLKEKVREKKEKVKMKEKEBEVTKACKACADKT	773
Db	527	ELHEKKRWASMEEQREYKVKRREBKKKLKEKAKER-----REKE-----	569
Qy	774	LATORLEERQKQOMILEEMKKTEDMCLTDHOPLPDPSRVPGLT-LPSGAFSDCLTIVE	832
Db	570	-----MLERLEKQ-----KRYEDELGT-KNLPAPRLVDITPEGLPNTLFGDVAMVVE	615
Qy	833	FLHSFGKVLGFDPAKOVPSGLVQEGLLCOGDSLGEVQDILLVRLLLKAALHDPGFPSPQS	892
Db	616	FLSCYGLLLPDAQYPIYAVS-LMEALADKGGFL-VLNRVLVLTLLQLODEAEDYGE-	672
Qy	893	LKILGEKVSIEPILRDNVNSILRCFLMAYG-----VXPALCD	929
Db	673	-----LGMKLSIPIULHSVSELVRLCLRRSDVQSESGSDTDNDKDSNAFPEDNEVQDFLE	729
Qy	930	RLRTQPQAQPPOOKAAVLAPFVHELNGSTLIINEIDKTLSSMYSRKNKWIVEGRLLRL	989
Db	730	KLETSEPFELTSEBKQILTALCHRI---LMTYSVQDHMETRQMSAELW--KERLAVL	783
Qy	990	K-----TVLAKRTGRSEVEM-----GRPECLGRRSSRIRMEETSGHBEESIESIAVPG	1040
Db	784	KEENDKKRABKQRKEMEAKNKGKVENGLGKTDKRKRIVKPEQPQVDTAEADMSIAVKS	843
Qy	1041	RR-----GRDGEVDATASSIPELEROIEKLSKR-----QLFPRKLLHSSQMLRAVSL	1089
Db	844	RLLLATQAKEREIQREREMKV-KLERQABEERIRKXKAAAEKAFQEGIAKAKLVMRTPI	902
Qy	1090	GODRYRRRYWVL-PYLAGIFVEGTEGNLVPVEVIKKETDSLKVAHAASLNPALFSMKMEL	1148
Db	903	GTDNRNHRNYLWLFSDVPGLFIE-----KGWVHDSID-----	933
Qy	1149	AGSNTTASSPARABSRPLTKPGFMQPRHFKS-PVRQDSQEQPAQLOLHPVAQPO	1207
Db	934	-----YRFNHHCXKHTVSGCEDYCPRSK-----	956
Qy	1208	PQLQLQLOKHGFLEQEGSPLSGQHDLSQAFLSWLSQTOHSGSLSSSVLTDPDSSP	1267
Db	957	-----KANLGN-----ASMNTQHGATGATEVAVETITPKQ	985
Qy	1268	GKLDPAQSPQPEPPEDAESSPDLOAFWNISIQMPCNAAPTPLAVSEDQTPSPQQL	1327
Db	986	G-----QNLWFLCDSQ-----	996
Qy	1328	ASSKPMNRPAAANPCSPVQSSSTPLAGLAPKRAGDPGEMPSQPTGLQPKRRGRPPSKF	1387
Db	997	-----	996
Qy	1388	FKOMEORYLTQTAQVPPPEMCGWWIIPDPEMLDAMLKALHPRGIREKALHKHLNK-HR	1446
Db	997	-----KELDELNLCHPQGIRESQLKERLEKRYQ	1025
Qy	1447	DFLOEVCILRSADPIFEPRQLPAPQEGIMSNSPKE-----KYETDVLAVLQWVEELEQR	1500
Db	1026	DIHSLHL--ARKPNLGLKSCDGNQB--LNLFLRSLDIEVATRLQKGLGVVEESTSFEAR	1082

```

RESULT 19
US-09-839-479-68
; Sequence 68, Application US/09839479
; Patent No. 6727222
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042002
; CURRENT APPLICATION NUMBER: US/09/839,479
; CURRENT FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 68
; LENGTH: 1525
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-839-479-68

```

Query Match	5.3%	Score 521.5;	DB 4;	Length 1525;
Best Local Similarity	19.6%;	Pred. No. 2.4e-31;		
Matches	281;	Conservative 185;	Mismatches 444;	Indels 523; Gaps 48;
Qy	565	EVIKYLSRLNLVHSVRREHSFSRPM-----VGDFEERTDTPEGLQWVQLSAEIPS	616	
Db	368	EMMKMSPNKLHT-----NPHIPKGGPPAKKPG---KHSKP-----LXAKGRSK	409	
Qy	617	RI---QAITGKRGRPR---NTEAKTKPKVKRGRGRPPKVKITEL-----LNKTDNR	664	
Db	410	GILNGQKSTGNSPKPKGLTKPTKMKQMTLLDMAKGTQ---KMTAPRNSGGTPTRTSSK	466	
Qy	665	PLKKLEA-----QETLNBEEDAKIAKSKKKQKVQVGRGCLTTIQGAANKRKQET	715	
Db	467	PHKHLPPAHLHLYAYYKENKDRDKRSALSCVTSKTARLLSSEDRARLPBELSLVQKRY	526	
Qy	716	KSLLHKPE--AKKKSXAEEKGKGTQKQBLKEKVXREKKEKVKMKKEEVTAKAPACKADKT	773	

Db 527 ELLEHKKRWASMEBQRKEYLKKREELKKLKEKAKER-----RKEK----- 569
Qy 774 LATQRLEERQKQMILEMKKPTEDMCLTDHQPDPFSRVPGLT-LPSGAFSDCLTIVE 832
Db 570 -----MLERLEKQ-----KRYEDQELTG-KNLPAPFLVDTPEGLPNTLFGDVAMVVE 615
Qy 833 FLHSGKVLGPDPAKDVPSLGVLOGLLCOGDSLGEVODLLVRLKAAALHDPGPPSYCQS 892
Db 616 FLSCYSGLLPDAQYPITAVS-LMEALADKGGFL-YLNRVLVILLQTLQDEIAEDYGE- 672
Qy 893 LKILGEKVEIPLTRDNVSEILRCFLMAYG-----VXPALCD 929
Db 673 ---LGMKLSIPLTHSVSELVRLCLRSDDVQESGSDTDNDKDSAAFEDNEVDPELE 729
Qy 930 RLRTOPPOAQPOQKAAVLAPFPVHELNGSTLIINEIDKTLESMSYRKNKWIIVGRRLRL 989
Db 730 KLETSFPELTSEKLIQILTALCHRI-----LMTYSVQDHMETRQOMSABLW--KERLAVL 783
Qy 990 K-----TVLAKRTGRSEVEM-----GRPECLGRRSSRIMEETSGMEBEESIAAVPG 1040
Db 784 KEENDKGAEKQKREMEAKNKENGKVGNGKTDKRIKRVKFBPQVDTEAEDMISAVKS 843
Qy 1041 RR-----GRRDGEVDATASSIPELERQIEKLSKR-----QLFFRKLLHSSOMLRAVSL 1089
Db 844 RLLAIQAKKEREIOEREMKV-KLEROAEEERIRKHKAAAEKAPQEGIAKAKLVNRTPI 902
Qy 1090 GODYRRRYWVL-PYLAGIFVEGTGNLVPBEVIKKTETDSLKVAHAHSLNPALEFSMKMEL 1148
Db 903 GTDRNHNRYLWFSDEVPLGFIET-----KGWVHDSID----- 933
Qy 1149 AGSNTTASSPARSRPLKTRFGFMQPRHFKS-PVRGQDSQPQAQLQPEAQLHVPAQPQ 1207
Db 934 -----YRFNHCKDHTVSGCEDYCPRSK----- 956
Qy 1208 PQLOLQLOSHKGFLQEGSPLSIGOSQHDLSQAFSLWSLQTSQSHSSLLSSVLTPDSSP 1267
Db 957 -----KANLKN-----ASMTQHGATYAVAVETTPKQ 985
Qy 1268 GKLDPAQSPPEPEPDEABESPDQAFWNISAQMPCNAAAPTPLAVSEDPQTPSPQOL 1327
Db 986 G-----QNLWFLCDSQ----- 996
Qy 1328 ASSKPMNRPAAANPCSPVOFSSTPLAGLAPKRRAGDPGEMPOSPTGLCQPKRGRPPSKF 1387
Db 997 ----- 996
Qy 1388 FKQMEQRYLTQTAQVPPEMCSGWWIIPDPEDMLDALKALHPRGIREKALHKLHK-HR 1446
Db 997 -----KELDELINCLHPQGIRESQLEKREKRYQ 1025
Qy 1447 DFLQEVCLRPADPIFEPRQLPAPFQEGTMSWPK-----KTYETDLAVLQWVEELQR 1500
Db 1026 DIIHSLHL--ARKPNLGLKSCDGNQE--LLNFLRSLDIEVATRLQKGLGYVEETSEFEAR 1082
Qy 1501 VIMSDLQTRGTCRSPDSTREDLAYCEHLSQEDITWRGPGREGLAPQRTYTNPLDLAV 1560
Db 1083 VIS-----LEKLKDFGECV----- 1096
Qy 1561 MRLAALQONKRRYLREPLMP-----THEVLEKALLSTPNGAPEGTTTTEI 1606
Db 1097 ---TALQASVTKKFLQGMFAPQKRRKLQSDSASKEVEDEKKWVE----- 1140
Qy 1607 SYEITPRIRIWRQTLQRCSAAHVCLIGHLERSIAWKSYNKVTCLVCRKGDNDEFULL 1666
Db 1141 EAKVASALEKWKTAIREAQTESRMHVLLGMLDACIKWDMSAENARCKVCPKKGDDDKLIL 1200
Qy 1667 CDGCDRGCHYIYCHRPKMEAVPEGDFWCTVCLAQOQVEGFTQKPGPKGQKRGKSYSLNF 1726
Db 1201 CDECNKAPHLFLCLRALYEVDPGEMQCPCAC-----OPATARNSSGRNYTBESA 1249
Qy 1727 SE-----GDGRRRRVLLKGRES-----PAA--GPR-----YS 1751

Db 1250 SEDSESDSEBEEEEEEEEEDYEVAGRLRLPRKTIRGKHSVIPPAAASRRRPGKPKPHS 1309
Qy 1752 EERLSPSK-----RRRLSMENHSHDLTFCEIILMEMESHDAAMPFXEPVNPRL 1799
Db 1310 TRRSQPKAPPVDAEVDLVTQTKSSRRSQSLELQKCEIILHKIVKYRFSWFPBSPVTRDE 1369
Qy 1800 VSGYRRRIKPNPMDSTMRERLLRGYTSSEBEFAADALLVFNDCOTFNEDDSEV 1852
Db 1370 AEDYDVTMPMDRQTVQNKSCSGSYRSVQEFLLDMKQVFTNAEYVNCRGS HV 1422
RESULT 20
US-09-418-710-27
; Sequence 27, Application US/09418710
; Patent No. 6596482
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTONAL REGULATOR
; FILE REFERENCE: 06501-042001
; CURRENT APPLICATION NUMBER: US/09/418,710
; CURRENT FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 27
; LENGTH: 1527
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-418-710-27
Query Match 5.3%; Score 521.5; DB 4; Length 1527;
Best Local Similarity 19.4%; Pred. No. 2.4e-31;
Matches 264; Conservative 179; Mismatches 440; Indels 479; Gaps 42;
Qy 619 QAITKGRGRPR--NTEKAKTKVEKVKRGRGRPPKVKITEL-----LNKTDNRPLKL 669
Db 415 QKSTGSKSPKGLKTPKTKMKQMTLLDMAKGTQ---KMTAPRNSGCTPRTSSKPKHL 471
Qy 670 EA-----QETLNEEDKAKIAKSKKWKQKQVORGECLTTIOGQARNKEKQETSKLH 720
Db 472 PPAALHLIAYKENKREDKRSALSCVISKTARLLSSEDRARLPEELSLVQKRYELLEH 531
Qy 721 KE--AKKKSAAEKGTQKQKLEKVKREKKEKVKMKKEEVTAKPACKADKTALQR 778
Db 532 KXRWASMEBQRKEYLKKREELKKLKEKAKER---REKE----- 569
Qy 779 RLEBRQKQMILEMKKPTEDMCLTDHQPDPFSRVPGLT-LPSGAFSDCLTIVEFLHSF 837
Db 570 MLERLEKQ-----KRYEDQELTG-KNLPAPFLVDTPEGLPNTLFGDVAMVVEFLSCY 620
Qy 838 GKVLGPDPAKDVPSLGVLOGLLCOGDSLGEVODLLVRLKAAALHDPGPPSYCQSLKILG 897
Db 621 SGLLLPDAQYPITAVS-LMEALSADKGGFLYLNRLVILLQTLQDEIAEDYGE---LG 675
Qy 898 EKVSEIPLTRDNVSEILRCFLMAYG-----VXPALCDRLRTQ 934
Db 676 MKLSEIPLTHSVSELVRLCLRSDDVQESGSDTDNDKDSAAFEDNEVDPELEKLETS 735
Qy 935 PFQAPQPOKAAVLAPFPVHELNGSTLIINEIDKTLESMSYRKNKWIIVGRLRLK---- 990
Db 736 EFPFLTSEKLIQILTALCHRI---LMTYSVQDHMETRQOMSABLW--KERLAVLKEEND 789
Qy 991 TVLAKRTGRSEVEM-----GRPECLGRRSSRIMEETSGMEBEESIAAVPGR--- 1042
Db 790 KRAEKQKREMEAKNKENGKVGNGKTDKRIKRVKFBPQVDTEAEDMISAVKSRLLA 849
Qy 1043 --GRDGEVDATASSIPELERQIEKLSKR-----QLFFRKLLHSSOMLRAVSLGDYR 1094

Db 850 IQAKKEREIOEREMKV-KLERQAEERIRKHAIAAEKAFQEGIAKAKLVMBRTPIGTORN 908
Qy 1095 RRRYWL-PYLAFIVFEGTEGNLVEEVIKETDSLKVAHAASLNPALFSMKMELAGSNT 1153
Db 909 HNRWYLFSDVPLGFIIE-----KGWVHDSID----- 934
Qy 1154 TASSPARARSPLKTKPGFMQPRHFKS-PVRGQDSEQQAOLQPEAQLHVPAPQPOLQL 1212
Db 935 -----YRFNHCKDHTVSGDEDYCPRSK----- 957
Qy 1213 QLQSHKGFLEQSGSLSGOSQHDLSQAFSLWSLQTSQSHSSLSLSSSVLTPDSSPGKLD 1272
Db 958 -----KANLGKN-----ASMTQHGCTATEVAVETTTPKQ----- 987
Qy 1273 APSOPPEPEDEAESPDLOAFWFNISAQPCNAAPTTPPLAVSBDQPTSPQQLASSKP 1332
Db 988 -----QNLWFLCDSQ----- 997
Qy 1333 MNRPSAANPCSPVQFSSTPLAGLAPKRRAGDPGEMPOSPTGLGQPKRRGRPPSPFKQME 1392
Db 998 ----- 997
Qy 1393 QRYLTQLTAQVPPPEMCGMWIIPDPEMLDAMLKALHPRGIREKALHKLANK-HRDFLOE 1451
Db 998 -----KELDELNLCHPOGIRESOQLERLEKRYODIHS 1031
Qy 1452 VCLRPSADPIPEPROLPAPQEGIMSPKE-----KTYETDLAVLQVVELEORVIMSD 1505
Db 1032 IHL--ARKPNLKLKSCDQNE--LNLFLRDLIEVATRLQKGLGYVEETSFEARVIS-- 1086
Qy 1506 LOIRWTCPSDPSTREDLAYCEHLSDSQEDITWRGPGREGLAPQRKTTNPLDLAVMLAA 1565
Db 1087 -----LEKLKDFGECV-----IA 1099
Qy 1566 LEQNVKRYLREPLW-----THEVVLEKALLSTPNAGPEGTTTISYBIT 1611
Db 1100 LOASVKKLFLQGFMAKQKRRKLOESDAKTEEVDEEKQWVE-----EAKVA 1146
Qy 1612 PRIIRWTLQRCRSAHVCLGLHLERSIAWEKSVNVTCLVCRKGDNDLFLLCDCGD 1671
Db 1147 SALEKWKTAIREAQTFSRMHVLGLMDACIKWDSAEANARCKVCPKGGEDDKLILCDECN 1206
Qy 1672 RGCHYCHRPKMEAVPEGDFCTVC-----LAQVGEFTQKPGPK 1713
Db 1207 KAFHLFLCLRPALYEPDGEWQCPACQATARRNSGRNYTBESASEDDDESDEEBEE 1266
Qy 1714 RGQKXKCYSLNFSGDCRRRLVLLKGRES--PAA--GPR-----YSEERLSPSK---- 1759
Db 1267 EEBEEDDYEV---AGLRPRKTIIRGKHSHVIPPAAARGRRPGKPHSTRSQKAPPVD 1323
Qy 1760 -----RRRLSMRNHSDLTFCIEILMEMESHDAAPFEPXEPVNPRLVSGYRRIKNP 1810
Db 1324 DAEVDLVLQYKRRSRROSLELOKCEELHLKIVKRFWSWPREPVRDEAREDDYDVITHP 1383
Qy 1811 MDFSTMRELLRGYTSSEFAADALLVDFNCQTFNEDDSEV 1852
Db 1384 MDFQTVQNKCSGYSRYSOEFLTDMKQVFTNAEVNCRGSHV 1425

RESULT 21

US-09-839-479-27
; Sequence 27, Application US/09839479
; Patent No. 6727222
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042002
; CURRENT APPLICATION NUMBER: US/09/839,479
; CURRENT FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JF98/01783
; PRIOR FILING DATE: 1998-04-17

; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 27
; LENGTH: 1527
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-839-479-27

Query Match 5.3%; Score 521.5; DB 4; Length 1527;
Best Local Similarity 19.4%; Pred. No. 2.4e-31;
Matches 264; Conservative 179; Mismatches 44; Indels 479; Gaps 42;

Qy 619 QAITKGRGRPR--NTEKAKTKVYKVRGGRGRPPKVKITEL-----LNKTDNRPLAKL 669
Db 415 QKSTGNSKSPKKGUKTKPTKMKQMTLLDMAKGTQ---KMTAPRNSGGTPTTSSKPHKHL 471
Qy 670 EA-----QETLNEEDKAKIAKSKKMKRQKVRGECCLTTIOQAARNKKEQETSKLKH 720
Db 472 PPAALHLIAYYKENKREDKESALSVCYSTARTLLSSEDRARLPEELRSLVQKRVELLEH 531
Qy 721 KE--AKKXSAEKXGKTQKBLKEKVKREKKEKVKMKKEEVTKAKPACADKTLATOR 778
Db 532 KKRWASMSSEORKEYLKKREELKKLKEKAKER---REKE----- 569
Qy 779 RLEERQKQMLLEMKKPTEDMCLTDHQPDPDFSRVPGLT-LPSGASDCLTIVEFLHSP 837
Db 570 MLERLEKQ-----KRYEDQELTG-KNLPAPFLVDTPEGLPNTLFGDVAMVVEFLSCY 620
Qy 838 GKVLGFPAPKVPISGLVQEGLLCQGSLSGVQDLLVRLKAAALHDPGFPSPYCSOSLKILG 897
Db 621 SGLLPDAQYPIITAVS-LMEALSADKGGFLVLRVLLQTLLODETAEDYGE---LG 675
Qy 898 EKVSEIPLTRDNVSEIILRCFLMAYG-----VXPALCDRLRTQ 934
Db 676 MKLSEIPLTHSVSELVRLCLRRSDVOEBSGSDTDNKKDSAAFEDNEVODEFLEKLETS 735
Qy 935 PFQAPQPOQKAAVLAFPVHELNGSTLIINEIDKLTLESMSYRKNKWIIVEGRLRLK--- 990
Db 736 EFFELTSEELKQILUTALCHRI-----LMTYSVDHMETRQMSAELW---KERLAVLEEND 789
Qy 991 TVLAKRTGRSEVEN-----GRPECLGRRSSRIMEETSGMEEEESIAAIVPGR--- 1042
Db 790 KKRACKQKREMEAKNKENGKVENGLKTDKRRKIVKPEPOVDTEAEDMIKAVKSRLLA 849
Qy 1043 --GRRDGEVDATASSIPELERQIEKLSKR-----QLFFRKLLHSSOMLRAVSLGQD 1094
Db 850 IQAKKEREIOEREMKV-KLERQAEERIRKHAIAAEKAFQEGIAKAKLVMBRTPIGTORN 908
Qy 1095 RRRYWL-PYLAFIVFEGTEGNLVEEVIKETDSLKVAHAASLNPALFSMKMELAGSNT 1153
Db 909 HNRWYLFSDVPLGFIIE-----KGWVHDSID----- 934
Qy 1154 TASSPARARSPLKTKPGFMQPRHFKS-PVRGQDSEQQAOLQPEAQLHVPAPQPOLQL 1212
Db 935 -----YRFNHCKDHTVSGDEDYCPRSK----- 957
Qy 1213 QLQSHKGFLEQSGSLSGOSQHDLSQAFSLWSLQTSQSHSSLSLSSSVLTPDSSPGKLD 1272
Db 958 -----KANLGKN-----ASMTQHGCTATEVAVETTTPKQ----- 987
Qy 1273 APSOPPEPEDEAESPDLOAFWFNISAQPCNAAPTTPPLAVSBDQPTSPQQLASSKP 1332
Db 988 -----QNLWFLCDSQ----- 997
Qy 1333 MNRPSAANPCSPVQFSSTPLAGLAPKRRAGDPGEMPOSPTGLGQPKRRGRPPSPFKQME 1392
Db 998 ----- 997
Qy 1393 QRYLTQLTAQVPPPEMCGMWIIPDPEMLDAMLKALHPRGIREKALHKLANK-HRDFLOE 1451

```
Db 998 -----KELDELLNCLHPQGIRSQERLKRKYQDIHS 1031
Qy 1452 VCLRPSADPIPEPRLQPAFQEGIMSWSPKE-----KTYETDLAVLQWVELEQRVIMSD 1505
Db 1032 IHL--ARKPNLGLKSCDGNQ--LLNFLRSDLIIEVATRLQKGLGYVEETSFEARVIS-- 1086
Qy 1506 LQIRGWTCPSPDSTREDLAYCEHLSDSOEDITWRGPGREGIAPQRKTTNPLDLAVMLAA 1565
Db 1087 -----LEKLDKDFGECV-----IA 1099
Qy 1566 LEQNVKRRYLREPLWP-----THEVYLEKALLSTPNGAPEGTTTISYBIT 1611
Db 1100 LQASVIKFLQGFMAPKQRRKLQSEDSAKTEEVDEEKKWVE-----BAKVA 1146
Qy 1612 PRIRIWRQTLORCRSAHVCLCLGHILERSIAWKSVMKVTCLVCRKGDNDFFLLDCGCD 1671
Db 1147 SALEKWKTAIREAQTFSRMHVLLGMLDACIKWMSAENARCKVCPKKGEDDKLILDCEN 1206
Qy 1672 RGCHLYCHRPKWEAVPEGDWFTVC-----LAQQVEGEFTOKPGPPK 1713
Db 1207 KAFHLFCURPALYEVDPGEWQCPACQATARNRGRNYTESASEDSEDESEEEEE 1266
Qy 1714 RGQKRKSGYSLNFSFGDRRRRLVLLKGRES---PAA--GPR-----YSEERLSPSK---- 1759
Db 1267 EEEEEEDYEV---AGLRPRKTIIRKXHSVIPPAARSGRRPGKKPHSTRSQPKAPPVD 1323
Qy 1760 -----RRRLSMRNHSDLTFCETIILMEMESHDAAPXEPVNPRLVSGYRRIKNP 1810
Db 1324 DAEVDELVLQKRSRRQSLQKCEILHKIVKYRFSWFPREPVTDRDEADYDVITHP 1383
Qy 1811 MDEFMRERLLRGVYTSSEEAADALVFDNCOTFNEDDSEV 1852
Db 1384 MDFQTVQNKCSGYSRVQEFELTMKQVTTNAEVNCRGSHV 1425

RESULT 22
US-09-418-710-29
; Sequence 29, Application US/09418710
; Patent No. 6596482
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042001
; CURRENT APPLICATION NUMBER: US/09/418,710
; CURRENT FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 29
; LENGTH: 1531
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-418-710-29

Query Match 5.2%; Score 512.5; DB 4; Length 1531;
Best Local Similarity 19.2%; Pred No.1.2e-30;
Matches 262; Conservative 179; Mismatches 446; Indels 475; Gaps 41;

Qy 619 QAITKRGKRP---NTEKAKTKVEPKVKRGRPPKVKITEL-----LNKTDNRPLKKL 669
Db 415 QKSTGNSKSPKGLKTPTKMKQMTLLDWAQGTQ---KWTAPRNSGGTPTRTSSKPHKHL 471
Qy 670 EA-----QETLNEDKAKIAKSKKKMRQKVORGECLTTIQOARNKKEQETKSLKH 720
Db 472 PPAALHLIAYYKENKDRKEDKFSALSCVISKTARLLSSSDRARLPEELRSLVQKRYELLEH 531
Qy 721 KE--AKKKSXAEEKGKTQEKLEKVKREKKEVKMKEKEVTKAKPACADKTLATOR 778
```

```
Db 532 KKRWASMSSEQRKEYLKKREELKKLKEKAKER---REKE----- 569
Qy 779 RLERQRQOQMLLENKXPTEDMCLTDHQPDPDFSRVPGLT--LPSGAPSDCLTIVVEFLHSP 837
Db 570 MLERLEKQ-----KRYEDQELTG--KNLPAFLVDTPEGLPNTFLFGDVAMVVEFLSCY 620
Qy 838 GKVLGFPAKDVPSLGVQLQGLLCCQDPSLGEVODLLVRLKAAALHDPGFSPYCSOSLKILG 897
Db 621 SGLLDPDAQYPTAVS--LMEALSADKGGFLVLRVLLVILLQTLLOLLOEIAIDYDGLG 679
Qy 898 EKVSEIPLTRDNVSEIILRCFLMAYG-----VXPALCDRLRTQ 934
Db 680 MKLSEIPLTLHSVSELVRLCLRRSDVQESGSDTDNDKDSAAAFEDNEVODEFLEKLETS 739
Qy 935 PFOAQPPQQAALVAFPVHELNGSTLIINEIDKTLBESMSSYRKWKWIVVEGBRLRLK---- 990
Db 740 EFFELTSEELQIILTALCHRI---LMTYSVQDHMETRQOMSABLW--KERLAVLKEEND 793
Qy 991 TVLAKRTGRSRVEM---GRPECLGRRSSRIEETSGMEEEEEESIAAIVGRR--- 1042
Db 794 KKRABKQKRKEMEAKNKENGKLVGLKTRKKRIVKFEPOVDTEAEDMISAVKSRLLA 853
Qy 1043 --GRRDGEVDATASSIPELERQIEKLSKR-----QLFRRKLLHSSQMLRAVSLGQDRY 1094
Db 854 IQAKKEREIQEREMKV--KLERQAEERIRKHAAAAKAFQEGIAKAKLVMRRTPIGTDRN 912
Qy 1095 RRRYVWL--PYLAGIFVGTGTEGNLVPEEVIKETDLSKVAHAASLNPAFLFSKMLAGSNT 1153
Db 913 HNRVWLFSDVPGULFIE-----KGWVHDSID----- 938
Qy 1154 TASSPARARSPLTKPGFMQPRHFKS--PVRGQDSEQPOAQLOPEAQLHVPAQPQLOL 1212
Db 939 -----YRFNHCKDHTVSGDEDYCFPSK----- 961
Qy 1213 QLQSHKGFLEQSGPLSGSQHDLQSASFLSWLSQTSQSHSSLLSSSVLTPDSSPGKLDP 1272
Db 962 -----KANLGKN-----ASMNTQHGTA TEAVETTPKQG--- 991
Qy 1273 APSQPEEPPEDEAESPDLOAFWNISAQWPCNAATPPPLAVSEDOPTSPQQLASSKP 1332
Db 992 -----QNLWFLCDSQ----- 1001
Qy 1333 MNRPSAANPCSPVQFSTPTPLAGLAPKRRAGDPGEMPOSPTGLQPKRRRPPSKFFKQME 1392
Db 1002 ----- 1001
Qy 1393 QRYLTQLTQAPVPPPMCGSMWMIIPDPEMLDAMLKALHPRGIREKALKHKLHK--HRDFLOE 1451
Db 1002 -----KELDELLNCLHPQGIRESQLEKREKRYQDIHS 1035
Qy 1452 VCLRPSADPIPEPRLQPAFQEGIMSWSPKE-----KTYETDLAVLQWVELEQRVIMSD 1505
Db 1036 IHL--ARKPNLGLKSCDGNQ--LLNFLRSDLIIEVATRLQKGLGYVEETSFEARVIS-- 1090
Qy 1506 LQIRGWTCPSPDSTREDLAYCEHLSDSOEDITWRGPGREGIAPQRKTTNPLDLAVMLAA 1565
Db 1091 -----LEKLDKDFGECV-----IA 1103
Qy 1566 LEQNVKRRYLREPLWP-----THEVYLEKALLSTPNGAPEGTTTISYBIT 1611
Db 1104 LQASVIKFLQGFMAPKQRRKLQSEDSAKTEEVDEEKKWVE-----BAKVA 1150
Qy 1612 PRIRIWRQTLORCRSAHVCLCLGHILERSIAWKSVMKVTCLVCRKGDNDFFLLDCGCD 1671
Db 1151 SALEKWKTAIREAQTFSRMHVLLGMLDACIKWMSAENARCKVCPKKGEDDKLILDCEN 1210
Qy 1672 RGCHLYCHRPKWEAVPEGDWFTVC-----LAQQVEGEFTOKPGPPK 1713
Db 1211 KAFHLFCURPALYEVDPGEWQCPACQATARNRGRNYTESASEDSEDESEEEEE 1270
Qy 1714 RGQKRKSGYSLNFSFGDRRRRLVLLKGRES---PAA--GPR-----YSEERLSPSK---- 1759
```


Db 1271 EEEEEEDYEV---AGLRPRKTKIRGKHSVPPAARSGRRPGKPKPHSTRSQKAPPVD 1327
Qy 1760 -----RRRLSMRNHSDLTFCFIIIMEMESHDAAPFEPVNPRLVSGYRIIRKNP 1810
Db 1328 DAEVDLVLQTKRSSRRQSLQCKEILHKIVKRFSPFPFVTRDEADYDVITHP 1387
Qy 1811 MDFSMTRELLRGYTSSEFAADALLVDFNCQTFNEDDSEV 1852
Db 1388 MDFQTVQNKSCGYSRVSQEFLLDMKQVFTNAEVNCRGSHV 1429

RESULT 23

US-03-839-479-29
; Sequence 29, Application US/09839479
; Patent No. 6727222
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042002
; CURRENT APPLICATION NUMBER: US/09/839,479
; CURRENT FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 29
; LENGTH: 1531
; TYPE: PRT
; ORGANISM: Homo sapiens
US-03-839-479-29

Query Match 5.2%; Score 512.5; DB 4; Length 1531;
Best Local Similarity 19.2%; Pred. No. 1.2e-30;
Matches 262; Conservative 179; Mismatches 446; Indels 475; Gaps 41;

Qy 619 QAITGKGRPR---NTEAKTKVEPKVGRGRPPKVKITEL-----LNKTDRPLKKL 669
Db 415 QKSTGNSKSPKGLTKPTKQMTLLDMWGTQ---RMTAPRNSGGTPTSTSKPHKL 471
Qy 670 EA-----QETLNEDDKAKIAKSKKMKQKQVQGECLTTIQQAENKQKQETKSLKH 720
Db 472 PPAALHLIAYKENKDRDKRSALSCVTSKTARLLSSEDRARLPEELSLVQKRYELLEH 531
Qy 721 KE--AKKXSAEKEGKTQKQKKEKVKREKKEKVEKKEEVTAKKPACKADKTLATOR 778
Db 532 KRWASMSSEQRKEYLKKKREELKKLKEKAKER---REKE-----REKE----- 569
Qy 779 RLERQKQOQMLEEMKKTEDMCLTDHQLPDPFSRVPLGT-LPSGAFSDCLTIVFELHSP 837
Db 570 MLERLEKQ-----KRYEDQELTG-KNLPAFLVDTPEGLPNTLFGDVAWVVEFLSCY 620
Qy 838 GKVLGFDPAKDVPSLGLVQELGLQCGDSIGVEQDLVLLKAAALHDPGFPYCSQSLKTLG 897
Db 621 SGLLPDQAYPITAVS-LMEALSADKGGFLYNRLVLLQTLLOLQDEIADYDYGELG 679
Qy 898 EKVEIPIPLTRDNVSEILRCFLMAYG-----VXPALCDRLRQ 934
Db 680 MKLSEIPLTLHVSSELVCLRRSDVQESSEGSTDDNKSAAAFEDNEVQDFLEKLETS 739
Qy 935 PFQAPPOQKAAVLAPFVHELNGSTLIINEIDKTLKESMSYRKNKWIIVEGRRLRLK---- 990
Db 740 EFFELTSEKLIQILTALCHRI-----LMTYSVDHMETRQOMSAELW--KERLAVLKEEND 793
Qy 991 TVLAKRTGRSEVEM-----GRPECLGRRRSRMEETSGNEEBEESIAAVGRR---- 1042
Db 794 KGRAEKQKRKEAMEAKENGKVGNGLTKDRKIRIVKPEQVDTEADMI SAVKSRLLA 853

RESULT 24

US-09-418-710-66
; Sequence 66, Application US/09418710
; Patent No. 6596482
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042001
; CURRENT APPLICATION NUMBER: US/09/418,710
; CURRENT FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17

Qy 1043 --GRRDGEVDATASSIPELEROIEKLSKR-----QLFFRKXKLHSSQMLRAVSGQDRY 1094
Db 854 IQAKKREIQSRMKV-KLERQAEERIRKHAAKAEKAFQEGIAKAKLVMRRTPIGTDN 912
Qy 1095 RRRYVWL-PYLAGIVFEGTEGNLPEEVIKKTOSLKVAHAASLNPALFSSMKWELAGSNT 1153
Db 913 HNRWLFSDVEVPLFIE-----KGWVHDSID----- 938
Qy 1154 TASSPARARSRLTKTKPGFMQPRHFKS-PVRGQDSEQPQALQPEAQLHVPAPQPOLQL 1212
Db 939 -----YRNNHCKOHTVSGDEDYCPRSK----- 961
Qy 1213 QLQSHKGLEQEGSPLSGOSQHDLSQSAFLSWLSQTSQSHSSLLSSSVLTTPDSSPGKLOP 1272
Db 962 -----KANLGKN-----ASMTQHGTAATEVAVETTPKQG---- 991
Qy 1273 APSQPEPEPEDEAESPDLOAFWNISAO MPCNAATPPLAVSEDOQTTPSQQLASSKP 1332
Db 992 -----QNLWFLCDSQ----- 1001
Qy 1333 MNRPSAANPCSPVQFSSTPLAGLAPKRRAGDPGEMPOSPTGLGQPKRRGRPPSKFKQME 1392
Db 1002 ----- 1001
Qy 1393 QRYLTQTLTAQVPPPEMCGMWWIIPDPEMLDAMLKALHPRGIREKALHKLHNK-HRDFLOE 1451
Db 1002 -----KELDELLNCLHPQIGRESQKLERLEKRYDIIHS 1035
Qy 1452 VCLPSADPIPEPRQLPAFOEGIMSWSPKE-----KTYETDLAVLOWVELEQRIVMSD 1505
Db 1036 IHL--ARKPNLGLKSCDGNQES-LLNLRSLDIIEVATRLQGGGLGYVEETSEFEARVIS-- 1090
Qy 1506 LQIRGWTCPSPDSTREDLAYCEHLSDSQEDITWRGPGREGLAPQKRTTNPLDLAVMLAA 1565
Db 1091 -----LEKLXDFGECV----- 1103
Qy 1566 LEQNVKRRYLRPLWP-----THEVVLEKALLSTNGAPEGTTTSEISYBIT 1611
Db 1104 LQASVKKFLQGFMAPKQKRRKLOSEDSAKTEEVDEEKKMVE-----EAKVA 1150
Qy 1612 PRIWQTQLORCSAAHVCLCLGHLERSIAWESVNVKVTCLVCRKGDNDRELLDCDGD 1671
Db 1151 SALEKWKTAIREAQTFSRMHVLLGLMDLACIKWDSAEANARCKVCPKKGEDKLLCDECN 1210
Qy 1672 RGCHYCHRPKMEAVPEGDNFCTVC-----LAQVVEGEFTOKPGPK 1713
Db 1211 KAFHLFCLRLPALYVPGEWQCPACQATARNRGRNYTEESASEDSEDESEEE 1270
Qy 1714 RGQRKSGYSINPSEGDRRRRLLKGRES---PAA--GPR-----YSEERLSPSK---- 1759
Db 1271 EEEEEEDYEV---AGLRPRKTKIRGKHSVPPAARSGRRPGKPKPHSTRSQKAPPVD 1327
Qy 1760 -----RRRLSMRNHSDLTFCFIIIMEMESHDAAPFEPVNPRLVSGYRIIRKNP 1810
Db 1328 DAEVDLVLQTKRSSRRQSLQCKEILHKIVKRFSPFPFVTRDEADYDVITHP 1387
Qy 1811 MDFSMTRELLRGYTSSEFAADALLVDFNCQTFNEDDSEV 1852
Db 1388 MDFQTVQNKSCGYSRVSQEFLLDMKQVFTNAEVNCRGSHV 1429

```
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 66
; LENGTH: 176
; TYPE: PRT
; ORGANISM: Homo sapiens
;
US-09-418-710-66
```

```
Query Match          4.8%; Score 472; DB 4; Length 176;
Best Local Similarity 54.5%; Pred. No. 5,4e-29;
Matches 96; Conservative 30; Mismatches 50; Indels 0; Gaps 0;

Qy 780 LEERKQOMILEMKKPTEDMCLTDHQPLDPFSRVPGTLTPSGAFSDCLTIVEFLHSGK 839
Db 1 LEQRLELEMAKELKKPNEDMCLADQKPLPRLIPGLVLSGSTFSDCLMVVQFLRNF GK 60

Qy 840 VLGEDPAKDVPSLGVLOEGLICQGSLSGEVDLLVRLKKAALHDPGPPSYCQSLKILGEK 899
Db 61 VLGEDVNTDVNLSVLQEGLLNIGDSMGVEVDLLVRLLSAAVCDPGLITGYKAKTALGEH 120

Qy 900 VSEIPLTRDYNVEILRCFLMAYGVKXPALCDRLRTQPPFOAQPPQKAAVLAPPVHEL 955
Db 121 LNVGVNRDYNVEILQIFMEAHCGQTETLSLTKAFQAHTPAQKASVLAFLINEL 176
```

```
RESULT 25
US-09-839-479-65
; Sequence 65, Application US/09839479
; Patent No. 6727222
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042002
; CURRENT APPLICATION NUMBER: US/09/839,479
; CURRENT FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 65
; LENGTH: 176
; TYPE: PRT
; ORGANISM: Homo sapiens
;
US-09-839-479-65
```

```
Query Match          4.8%; Score 472; DB 4; Length 176;
Best Local Similarity 54.5%; Pred. No. 5,4e-29;
Matches 96; Conservative 30; Mismatches 50; Indels 0; Gaps 0;

Qy 780 LEERKQOMILEMKKPTEDMCLTDHQPLDPFSRVPGTLTPSGAFSDCLTIVEFLHSGK 839
Db 1 LEQRLELEMAKELKKPNEDMCLADQKPLPRLIPGLVLSGSTFSDCLMVVQFLRNF GK 60

Qy 840 VLGEDPAKDVPSLGVLOEGLICQGSLSGEVDLLVRLKKAALHDPGPPSYCQSLKILGEK 899
Db 61 VLGEDVNTDVNLSVLQEGLLNIGDSMGVEVDLLVRLLSAAVCDPGLITGYKAKTALGEH 120

Qy 900 VSEIPLTRDYNVEILRCFLMAYGVKXPALCDRLRTQPPFOAQPPQKAAVLAPPVHEL 955
Db 121 LNVGVNRDYNVEILQIFMEAHCGQTETLSLTKAFQAHTPAQKASVLAFLINEL 176
```

RESULT 26

```
US-09-698-295-1
; Sequence 1, Application US/09698295
; Patent No. 6689584
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATORY FACTOR
; FILE REFERENCE: 06501-068001
; CURRENT APPLICATION NUMBER: US/09/698,295
; CURRENT FILING DATE: 2000-10-27
; PRIOR APPLICATION NUMBER: PCT/JP99/02340
; PRIOR FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: JAPAN 10/137631
; PRIOR FILING DATE: 1998-04-30
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 2907
; TYPE: PRT
; ORGANISM: Homo sapiens
;
US-09-698-295-1
```

```
Query Match          4.2%; Score 413.5; DB 4; Length 2907;
Best Local Similarity 18.5%; Pred. No. 2,2e-22;
Matches 375; Conservative 297; Mismatches 771; Indels 579; Gaps 83;

Qy 79 SSSSTSHLHPSVAYDCLWNYSQYPSANPGSNLKDPLLQSFGQYPLNGILGSRQPSS 138
Db 1234 SSSSAUHH-----SSVPKS---TNDRATPLSR-----AMDFEGLGCDSESN 1274

Qy 139 PSHTNLRAGSQKFWANGTHSPMGLNFDQSOLYDFPDQNFEEVCSGIIHPDEAAEKEMTS 198
Db 1275 TLENSDVTISIQ-----DSSEEDMIVQNSNESISISQFRTRE-QDV-----EVLEPLKE 1322

Qy 199 VVAENGTLVCSLELEXQPELKMCGYNGSVPSVESLHQEVSVLPDPTVSCLDPPSHLP 258
Db 1323 LVSGESTG-----NCEDRLP-VKGTGANGKKPSQOKLEE-----RPVVKCSQD----- 1365

Qy 259 DQLEDTPILSEDSLEPENSLAPEVPVSGGLYGIDDTLMGAEDKLP-----LXDS 307
Db 1366 -----IKUNTTDKNNENRESEKKGQR--TSFQINGKDNPKPKYLKGECLKEISE 1416

Qy 308 PVISALDCPSLNNATAFSLADDQSOTSTISIPASPTSPVLGESVLQD-----NSFD-----L 359
Db 1417 RVVSGNVPEPKVNNIN--KIIPENDIKSLTVKESAIRPFINGDVIMEDFNERNSETKSHL 1474

Qy 360 NNGSDAE---QEWE-----TQSSDFPESLTOPA--PQOSSTIQLHP---ATSPAVSPPTS 407
Db 1475 LSSSDAEGNYRDSLETLPTSKESDSTQTTTPSACSPESNYSNVQEDMEIETSEVKKVTSS 1534

Qy 408 PAVSLVVSPPASPEIS-----PEVCPAASTVW--SPAVFSVVS 443
Db 1535 PITSEESNLSNDFIDENGLPIKNENNVNGESKKKTVITEVTTTSTVATESKTVIKVEK 1594

Qy 444 PASSAVLPAVSLEVLPTASVTSKASPVTSKASPTSPAAAFTASPANKDVSSFLETTADVEETG 503
Db 1595 GDKQTVSSSTENCAKSTVTTTITVTKLSIPSTGSDVDIISVKESQSTVTVTTTVDLSLT 1654

Qy 504 EGLTASSGDVMRRRIATPEEVRLPLOHGWRREVRIRKKGSHRWQGETWYWGPGCKRMKQF 563
Db 1655 TGGTLVTSMTV-SKEYSTRDKVKL---MKFSRPKKTRSGT-----AL 1692

Qy 564 PEVIKYLSRNLVHSV-----RRR--HFSP-----SPRMPYGD 593
Db 1693 PSYRKFTVTSKTSKISFVLPNDCLKLARKGIRVFPFNYNNAKALDILWPSPRPTGI 1752

Qy 594 FFEER-DTPEGLQ-----WYQLSAEELPSRQIAITKRGRRPNTEKAKTKEVPKVK 643
Db 1753 TWRYRLQTVKSLAGVSLMLRLWLWASLRWDDMAAKVP---PGGSTRITSET----- 1801

Qy 644 RGRGPPVKITELLNKTDNRP-----LKKLEAQETLNEEDKAKIAKSKKQKQVK 695
Db 1802 -----EITTTTEIKRRDVGYPGIRFEYCIKRIICPIGVETPKETPTPQKGLRSSAL 1854
```

QY 696 RGECLTTTQOARNRKQKETSLSKHKEAKKXSAEKEKTKQKQLKQKVRKKEKVKVM 755
DB 1855 R-----PKRPETPKQTGPVITETWAAEELELMEIRAFARVEKEKAQAVEQ 1901
QY 756 KEKEVTKAKPACKADKTLATQRRLEERQKQOMILEEMKPKPTEDMCLTDHQPDPFSRVP 815
DB 1902 QAKKEL-----EQKPTVIATSTSTSTST-----1929
QY 816 GLTSPGAFDCLTIVEFLHSEKVLGPDPAKDVPSLGVLOBGLLCQSDLSGEVDLLVR 875
DB 1930 -----ISPAQKVMVAPI--SGSVTTGTGM-----1951
QY 876 LLKAALHDPGPSPYCSQSLKILGEKVEIPLTRDNVSEILRCLFMAYGVXPAALCDRLRTP 935
DB 1952 VLTTRVGPSPATVTFQONKFNHQTATVWKQGSN--SGVVQVQKQKVLGIPSTGTSTQ--QT 2009
QY 936 FOAQPPQKAAVLAFVPHLNGSTLIINEIDKLTLESMSYKRWKWIVEGRLLR-----LK 990
DB 2010 FTSFOP--RTATVTIRPNTSGSGT-----TSNSQVITGPQIRPQMTVIR 2052
QY 991 TVLAKRT--GRS-----EVEGRPECLGRRRSSRIMBETSGMBEEEEEEETAAVPGRR-- 1042
DB 2053 TPLQOSTLIGKAIIRPVMVQPAQOVM-----TQIRGQPVSTAVSAPNTVSSTPGQKS 2107
QY 1043 ---GRDGEVDATASSIPELERQIEKLSKQQLFFRKLLHSSQMLRAVSLGQDRYRRYV 1099
DB 2108 LTSATSTNIOSSASQPRPQGVQKLTMAQLTQLTQGHGNGQLTVVIQGGQTTGQLQ 2167
QY 1100 VLPYLAGIFVEGTEG-----LVPEVIKK-----ETDSLKVAHAASL 1137
DB 2168 LIP--QGVTVLPVPGQQQLMQAAMPNGTVORFLTPLATTATTSTTTTSTTAAGTGEQ 2225
QY 1138 NPALFSMMELAGSTNTTASSPARARS--RPLTKPKGFMQPRHFKSP--VRGQDSEQPPQAQL 1195
DB 2226 RQSKLSQMVQVHQDKTL--PPAQSSVGPAPAKQAPQAPQAPQAPQAPQAPQAPQAPQ 2283
QY 1196 PEAQL-----HVPAPQPOLQLOLQSHKGFLEQEGSPLSLGQSQHDLS--QSAFLSWLSQ 1248
DB 2284 PEVOTQTTVSHVSEAPQ--THAQSSKQVAAQSQPSNVQGSQVVRVQSP-----SQ 2335
QY 1249 TQSHSLSSSVLTPDS--SPG---KLDPAQAP--PEEPEPDEAESPDQAFWFNISQOM 1303
DB 2336 TRIRPS-----TPQSLSPGQSQVQTTTQPIPIQP-----HTSLQI 2372
QY 1304 PCNAAPTPLAVSEDPQTPSPQOLASSKPMNRPSAANPCSP--VQPSSTPLAGLAPKRRAG 1362
DB 2373 PSQGPQ-----SQVQVSSQTLSLGGTLNQVSSPSRPLQI-----2412
QY 1363 DPGEMQSPQGLGQPKRRGRPPSKPFQWMEQRYLTQLTAPVPPBMCSGMWIIPDPEMLD 1422
DB 2413 ---QOPQ--QVIAVPQLQQQ--QVLSQIQSQVVAQIQAOQ-----SG---VPQOIQLQ 2457
QY 1423 AMLKALHPRGIREKALHKLHNRDPLQEVCLRPSADPIFPRQLPAPQEGIMSWSPKEK 1482
DB 2458 LPIQIQSSAVQTHQIONVTVQAASVQB-----QLQRVQQ--LRDQQQKKK 2502
QY 1483 TVETDLAV-----LOWVEELEORVIM-----SDLOIRGWTCPSPDSTREDLAVC--- 1526
DB 2503 QQQIEINVTNPSKLIKIVETIQKQVVMKHNAVIEHLKQKSMTPAEREENQRMIVCNQVM 2562
QY 1527 ---EHLSDSQEDITWRGPGREGLAPQRKTNPLDLAVMLAALBQ-----NVKRRYL--- 1575
DB 2563 KYILDKIDKEEQAAKREESVQKRSQWATKLSALLFKHQLRAELIKKRALLDK 2622
QY 1576 -----REPLWPTHVLEKALLS-----TPNGAPEGTTTE 1605
DB 2623 DLQIEVQBELKRDLIKKEKOLMLQAQATAVAACPVPVTPVLPAPPAPPPPPPPPGVQH 2682
QY 1606 ISEYITPRIRIWRQTLOR-----CRSAAHVCLCLGLHLSIAWEKSNVKTCLVCRKG 1658
DB 2683 TGLISTPLPVASQKREBEKDSKSKKKMMISTTSKE-----TKDQTKLYCLICKTPY 2737
QY 1659 DNDEFLLLCDGCDRGCHICY---HRPKMEAVPEGDFWCTVCLAQQVGEFTQKGFPPKRG 1715

RESULT 27

US-09-698-295-10
; Sequence 10, Application US/09698295
; Patent No. 6689584
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATORY FACTOR
; FILE REFERENCE: 06501-068001
; CURRENT APPLICATION NUMBER: US/09/698,295
; PRIOR FILING DATE: 2000-10-27
; PRIOR APPLICATION NUMBER: PCT/JP99/02340
; PRIOR FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: JAPAN 10/137631
; PRIOR FILING DATE: 1998-04-30
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 2781
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-698-295-10

Query Match 4.1%; Score 406.5; DB 4; Length 2781;
Best Local Similarity 18.3%; Pred. No. 7,3e-22;
Matches 370; Conservative 298; Mismatches 775; Indels 579; Gaps 82;

QY 79 SSSTSHLHPSVAYDCLWNYSQYPSANPGSNLKPPLLSQFSGGQYPLNGILGSRQSPSS 138
DB 1108 SSSKALH-----SSVPKS---TNDRTATPLSR---AMDPEGLKGCDSSENS 1148
QY 139 PSHTNLRAGSOKFWANGTHSPMGLNFDSDQSLYSPDPONFEYVCSGIHPDEAAKEWTS 198
DB 1149 TLENSDTSVSIQ-----DSSEEDMIVQNSNESISEQFRTRE--ODV-----EVLPLKCE 1196
QY 199 VVAENGTVGLVCSLELEEXQPELKMCGYNGSVPSVSHQEVSVLVPDPTVSCLDPPSHLP 258
DB 1197 LVSGESTG-----NCEDRLP--VKGTEANGKKPSQOKKLEE-----RPVNCSDQ----- 1239
QY 259 DOLEDPTPILSBDLSLEPNLSLAPFVPSVGLYGDIDPTELMGAEDKLP-----LXDS 307
DB 1240 -----IKLKNVTDKKNENRESEKQGR--TSTFQINGKDNKPKVLKGBCLKEISES 1290
QY 308 PVISALDCPSLNNATATSLADDSTQSTISFASPTSPVPLGESVLQD-----NSPD-----L 359
DB 1291 RVWSGNVEPKVNNNTN--KIIPENDIKSLTVKESAIRPFINGDVMEDFNERNSSSETKSHL 1348
QY 360 NNGSDAE--OQEEME---TQSDFFPPSLTPQA--PDOSSTIQLHP---ATSPAVSPPTS 407
DB 1349 LSSDAEGNYRDSLETLPSTKESSTQTTTSPASCPSNSVNVQEDMEIETSEVKKVTSS 1408
QY 408 PAVSLVWSPAASPRIS-----PEVCPAASTTV--SPA VFSVVS 443
DB 1409 PITSEESNLNDFIDENGLPINKNENVNGESKKTIVITEVTTMTSTVATESKTIVKEK 1468
QY 444 PASSAVLPVSLVPLTASVTSPPKASPVTSFAAAPPPTASPAKOVSSLETADVEETIG 503
DB 1469 GDKQTWSSSTENCAKSTVTTTTTTTTLKSTPSTGSGVDIISVKQSKTVVTTTVDLSLT 1528

```
QY 504 EGLTASGSDVMRRRIATPEEVRLPLQHGWRREVRIKKGSHRWQGETWYGPCGKRMKQF 563
Db 1529 TGGTLVTSMTV-SKEYSTRDKVKL---MKFSRPKTKRSTG-----AL 1566
QY 564 PEVIKYLNRNLVHSV-----RRE--HFSF-----SPRMPYGD 593
Db 1567 PSYKFFVTKSTKSI FVLPNDDLKKLARKGGIREVPYFNNAKPAIDLWPPSPRPTGI 1626
QY 594 FFEER-DTPEGILQ-----WVLSABEIPSRIOAITGKRGRPRNTEKAKTEKVPVKV 643
Db 1627 TWRYRLQTVKSLAGVSLMLLLWASLRWDDMAAKVP-----PGGSTRTETSET----- 1675
QY 644 RGRGPPKVKITELANKTDNRP-----LKKLEAQETLNEEDKAKTAKSKKWRQKVQ 695
Db 1676 -----EITTEIIRKRDVGPYIRFYCIKKIICPIGVPTPKETPTPORKGLRSSAL 1728
QY 696 RGECLTTTQGOARNRKQETKSLKHKEAKKXAEKXGKTQKQKLEKVKREKKEKVKM 755
Db 1729 R-----PKRPETPKQTGPVIIETWABEELELWEIRAFARVERKEKAQAVEQ 1775
QY 756 KEKEVTKAKPACAKDKTLATQRRLEBRQKQOMILEEMKKPTEDMCLTDHQPDPDFSRVP 815
Db 1776 QAKKRL-----EQCKPTVIATSTTSSTTST----- 1803
QY 816 GLTLPAGAFSDCLTIVEFLHSPGKVLGFDPAKDVPSLGVLOEGLL-CQGSLSGEVQDLLVR 875
Db 1804 -----ISPAQKQWVAPI--SGSVTTTKM----- 1825
QY 876 LLKAALHDPGFPSYCSKILGEKVEIPLTRDNVSEILRCFLMAYGVXKPCALCDRLRTPQ 935
Db 1826 VLTTKVGSFPAITVQONKNFHOTFATVWKVQGSN-SGVVQVQKVLGIIPSTGTSQ-QT 1883
QY 936 FOADPPQOKAAVLAPPVHELNGSTLIINEIDKTESMSSYRKNKWIVEGRLLR-----LK 990
Db 1884 FTSFOP-RTATVTRPNTSGSGT-----TSNSQVITGPQIRPGMTVIR 1926
QY 991 TVLAKRT-GRS-----EVENGRPEECLGRRSSRIMEETSGMEEBEESIAAVPGR- 1042
Db 1927 TPLQOSTLGKAIIRTPVMVQGAPOQVM-----TQIRGQPVSTAVSAPNTVSTSPGOKS 1981
QY 1043 ---GRRDEVDATASSIPELERQIEKLSKQLFRFKLLHSSQMLRAVSLGQDRYRRYV 1099
Db 1982 LTSATSTNIOSSASOPRPOGOVKLTMAQLTOLTQCHGNGOGLTVVIQGGQTTGQLQ 2041
QY 1100 VLPYLAGIFVSGTEGN-----LVPEEVIK-----ETDSLKVAHAASL 1137
Db 2042 LIP--QGVTVLPFGQQLMQAAMPNGTVQRELFPTLATTATTATTTTSTTAAAGTGEQ 2099
QY 1138 NPALFSMKWELAGSNVTTASSPARARS-RPLKTKGFMQPRHFKSP-VRGQDSEQPOAQLQ 1195
Db 2100 ROSKLSUPQMVHQDKTL---PPAQSSSVGPAKAQPOTAQPSARPQOTQPSAPAEVQTQ 2157
QY 1196 PEAQL-----HVPAAQPOLQLOLQSHKGFLQEGSPLSLQSQSHDLS-QSAFLSWLSQ 1248
Db 2158 PEVQTQTTVSHVPEAPF---THAQSSKPVQAQSQPSQSNVQGSPPRVQSP-----SQ 2209
QY 1249 TQSHSLSSLLSSVLTPTS-SPG---KLDPAQSP-PEEPEDEAESPPDQAFWNISQM 1303
Db 2210 TRIIRPS-----TPSQLSPGQSQSVQTTTSPQPIQP-----HTSLQI 2246
QY 1304 PCNAAPTPLAVSEDOPTPSQQALSSKPMNRPSAANCSP-VQFSSTPLAGLAPKBRAG 1362
Db 2247 PSQGOPO-----SQPVQSSQTLSGGQTLNQVSVSPSRPOLQI----- 2286
QY 1363 DPGEMPSQPTGLGQPKRGRPPSPFKQMEQRYLTQLTAQVPPPMCSGMMWIPDPEMLD 1422
Db 2287 ---QOPO-POVIATVPOLQOQ--VQVLSQIQSVQAQIAQO-----SG---VQOQIKLO 2331
QY 1423 AMLKALHPRGIREKALHKLHKNHRDFLQEVCLRPSADPIFEPRQLPAPQEGIMSNSPKEK 1482
Db 2332 LPIQIQSSSAVQTHQIQNVVTVQAASVQE-----QIQRVQO-LRDOQQKKK 2376
```

```
QY 1483 TYETDL-----AVLQWVEELQORVIMSDLOIRGWTCPSPDSTREDLAYC----- 1526
Db 2377 QOOIEIKREHTLOASNOSEIIOQVVMKHNAVIEHLKQKKSMTFAEREENORMIVCNQVM 2436
QY 1527 ---EHLSDSOEDITWRPGREGIAPORKTTNPDLDLAVMLRAALEQ-----NVKRRYL--- 1575
Db 2437 KYILDKIDKEKQAAKKRKEESVEQKRSQKQATKLSALLFKHKEQLRAEILKKRALLDK 2496
QY 1576 -----REPLWPTHVEVLEKALLS-----TPNGAPEGTTTE 1605
Db 2497 DLQIEVOBELKRDLIKKEKDLMLQLAQATAVAAPCPVTPVLPAPPAPPPPPPPGVQH 2556
QY 1606 ISYEITPRIRWROTLOK-----CRSAAHVCLCLGHLERSIAWESVNVKTVCLVCRKG 1658
Db 2557 TGLLSTTTLPVASOKRKEEBEKSSSKKKMMISTTSKE-----TKDXTLYCLICKTPY 2611
QY 1659 DNDEFLLDCDCGRGCHIYC---HRPKMAEVPEDGWFCTVCLAOQVEGEFTQKPGFKRG 1715
Db 2612 DESKFYIGCDRCQNWYHGRCVGILQSEAEILDE--YVCPQC--QSTEDAMTV----- 2659
QY 1716 QKRXSGYSLNPFSEGDGRRRVLKGRSPAAGPRYSEERLSPSKRRRLSMENHSHDLTFC 1775
Db 2660 -----LTPLTEK-----DYBGL 2671
QY 1776 EIIIMEMESHDAWPFPEPNRPLVSGYRRIIKNPMDFPMRERLLRGYTSSEFAADA 1835
Db 2672 KRVLRSLQAHKMAWPFLEVDNDAPDYGVIKFPMDLATMEERVQRYYEKLTFFVADM 2731
QY 1836 LLVFDNCQTNEDDSEVGKAGHIMRRFPESRWEEFYQKQAN 1877
Db 2732 TKIFDNCRYNPSDSPFVQCAEVLESFFVKLGKFKASRSHN 2773

RESULT 28
US-09-949-016-10125
; Sequence 10125, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 10125
; LENGTH: 1155
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-10125

Query Match 3.4%; Score 335; DB 4; Length 1155;
Best Local Similarity 19.5%; Pred. No. 7.3e-17;
Matches 241; Conservative 148; Mismatches 449; Indels 400; Gaps 52;

QY 766 PACKADKTLATQRRLEBRQKQOMILEEMKKPTEDMCLTDHQPDPDFSRVPGTLPSGAFS 825
Db 146 PASLLDTCVQCQSQSQRREAB---PKLLPCLH8FCL---KCLPEPERQLSVPIPGSGNG 199
QY 826 D-----CLTIVEFLHSGKVLGFDPAKDVPSLGVLOEGLLCOQ--DSLGE 868
Db 200 DIOQGVIRCPVCQRCRQIDLVNRYFK-----DTSEAPSSSDEKSEQVCTSCEDNAA 254
QY 869 VQDILLVRLKAAALHDPGFPSYCSQ--LKLGEKVEIPLTRD-----NVSEILRCFLM 919
Db 255 V-----GFCVECGEWLCKTCTIEAQRVKFTKDLIRKEDVSESVG----- 295
```

```
QY 920 AYGVXPALCDRLRTPQFQAPQPOQAALVAPVHVLNGSTLIINEIDK-----TLESM 972
Db 296 ASGORPVFC-----PVHKQEQKLFCTCDRLTCDRCQQLLHKK 333
QY 973 -----SSYKKNKWIVEGRRLRLKTVLAKRT-----GRSEVEMGRPEECGLRRSSRIM 1020
Db 334 EHRVQFLAEAFQNGKAIENLLAKL-----LEKNYVHFAATQVQ-----NRIK 377
QY 1021 E--ETSGMEEBEEESIAAVPGRGRDRGEVDATASSIPELEROIEKLSKRLQFRRKKLL 1078
Db 378 EVNETNKRVEQEIKAIVFTL-----INEINKGKS---LQQLENVTKER---QMELL 424
QY 1079 HSSQMLRAVSLGQDRYRRYVWLPYLAGIFVETGEGNLVPEE-----VIKKTOSL 1129
Db 425 QQONDITGLS-----RQVKHVMFTNWAIASGSSSTALLYSKRLITFQLRHILKARCDPV 478
QY 1130 KVAHAH---SLNPALFSMKMELAGSNTTASSPARARSPLTKPGF----- 1172
Db 479 PAANGAIRFHCDPTFWAKNVNVLGNLVIESKPA-----PGYTNVVGVQVPPGTN 528
QY 1173 -----MOPRHFKSPVRGQDSEQQAOLQPEAQLHVAPOQOLQLOQSHKG 1219
Db 529 HISKTPGQINLAQLRQHQVVAQKHQLOQ-----QMRMQQPPAPVPTTTTITQOH-- 581
QY 1220 FLEQEGSPLSLGQ-----SQHDLQSASFSLWSLQTSQSHSSLSLSSVLTDPSSPG 1268
Db 582 --PRQAAPQMLQQOQPRLLISVQTMQRGNMNCAGF-----QAHQMLRAQNA--RIPG 629
QY 1269 KLDPAPOPPEPEPEDEAESPDLOAFWFNISAOQPCNAATPPPLAVSEDOPTSPQOLA 1328
Db 630 -----IPRHSGFOYSMMQPHLQROHNS-----PGHAGFPFVVSVHNTTINPTPTTA 676
QY 1329 SSKPMNRPSAANPCSPVQFSTPLAGLAPKRRAGDPGEMPOSPTGLGQPKRRGRPP----- 1384
Db 677 TWANANR---GPTSP-----SVTALSLPS--VTNPENLPSLP-----DIPPIQLE 717
QY 1385 ---SKFFKOMBORYLTQTATQVPPVEMCGMWIIPDEMLDAMLKALHPRGIREKALHKH 1441
Db 718 DAGSSSLDNLLSRY---ISGSHLPQPTSTWNPSPGPSALSPGSSGLS----- 762
QY 1442 LNKHRDFLQEVCLRSADPIEPEROLPAFOEGIMSWPK--EKT---YETDLAVLOWVEE 1496
Db 763 -NSH-----TVRPPTSTSGSRGSCSGRTAEKTSLSFKSDQVKVQBPBG 808
QY 1497 LEQRVIMSDQIRGWTCPSPDSTREDLAYCEHLSQSDIETWRGPGREGLAPQRTTNPL 1556
Db 809 TEDEI-----CSFSGGVKQ-----EKTEDGRSACMLSPSSSLTPPLSTNHL 852
QY 1557 DLAVMLAALAEONVKRRYLREPLWPTHEVLEKALLSTPNGAPEGTTTTEISYEITPRIRI 1616
Db 853 ESELDALASLENHVK---TEPA--DMNESCKQSGLSLNVNGK----- 889
QY 1617 WRQTLQRCRAHVCLCLGHILERSIAWEKSVNKKVTCLVCRKGDNDDEFLLCDGCDRGCHI 1676
Db 890 -----SPTRSLMHSRARIQGGDNN--KDDDPNEDCAVCONGGD---LLCCEKCPKVFHL 939
QY 1677 YCHRPKMAEVPBGDFWCTVCLAAQOQVEGFTQKPGFPKRGKRGKSGYSLNFSFEGDGRRRV 1736
Db 940 TCHVPTLLSFFSGDMICITFC----- 959
QY 1737 LLKGRSPAAGPRYSEERLSFSKRRRLSMRNHSDLTFCIILMEMESHDAAPFPXEPVN 1796
Db 960 ---RDICKPEVEYDCDNLQHSKKGKTAQGLSPVDORCERLLLYLCHLSIEFQEPV- 1014
QY 1797 PRLVSGYRIIKPNMDFSTMERLLR---GGYTSSEBFAADALLVDFNCQTFNE----- 1847
Db 1015 PASIPNYKIIKKPMDLSTVKKLOKHSQHYQIPDDFVADVRLIFKNCERFENMKVQV 1074
QY 1848 -----DDSVGKAGHIMRFFESRWEEFYQKG 1874
Db 1075 VYADQTEINLKADSEVAQGAVALYFEDKLTETIYSDR 1112
```

RESULT 29

```
US-09-949-016-10126
; Sequence 10126, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIORITY FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10126
; LENGTH: 1155
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-10126
```

Query Match 3.4%; Score 335; DB 4; Length 1155;

Best Local Similarity 19.5%; Pred. No. 7.3e-17;

Matches 241; Conservative 148; Mismatches 449; Indels 400; Gaps 52;

```
QY 766 PACKADKTLATQRLERQKQOIMLEMKKTEMCDMLTDHQPDPFDSRVPGLTLPDSGAFS 825
Db 146 PASLLDTCVQCQOSQSRREAE---PKLLPLCLHSFCL---RCLPEPERQLSVPIPGSNG 199
QY 826 D-----CLTIVEFLHSFGKVLGDFPAKDVPSLGLVQELGCG--DSLGE 868
Db 200 DIQGVGIRCVQCQECRQIDLVNDFVK-----DTSEAPSSSEKSEQVCTSCEDNASE 254
QY 869 VQDLLVRLKAAALHDPGPPSYCQS--LKILGEKVEISEPLTRD-----NVSEILRCFLM 919
Db 255 V-----GFCVECGEWLCKTCEAHQVRKFTKDLIRKKEDVSESVG--- 295
QY 920 AYGVXPALCDRLRTPQFQAPQPOQAALVAPVHVLNGSTLIINEIDK-----TLESM 972
Db 296 ASGORPVFC-----PVHKQEQKLFCTCDRLTCDRCQQLLHKK 333
QY 973 -----SSYKKNKWIVEGRRLRLKTVLAKRT-----GRSEVEMGRPEECGLRRSSRIM 1020
Db 334 EHRVQFLAEAFQNGKAIENLLAKL-----LEKNYVHFAATQVQ-----NRIK 377
QY 1021 E--ETSGMEEBEEESIAAVPGRGRDRGEVDATASSIPELEROIEKLSKRLQFRRKKLL 1078
Db 378 EVNETNKRVEQEIKAIVFTL-----INEINKGKS---LQQLENVTKER---QMELL 424
QY 1079 HSSQMLRAVSLGQDRYRRYVWLPYLAGIFVETGEGNLVPEE-----VIKKTOSL 1129
Db 425 QQONDITGLS-----RQVKHVMFTNWAIASGSSSTALLYSKRLITFQLRHILKARCDPV 478
QY 1130 KVAHAH---SLNPALFSMKMELAGSNTTASSPARARSPLTKPGF----- 1172
Db 479 PAANGAIRFHCDPTFWAKNVNVLGNLVIESKPA-----PGYTNVVGVQVPPGTN 528
QY 1173 -----MOPRHFKSPVRGQDSEQQAOLQPEAQLHVAPOQOLQLOQSHKG 1219
Db 529 HISKTPGQINLAQLRQHQVVAQKHQLOQ-----QMRMQQPPAPVPTTTTITQOH-- 581
QY 1220 FLEQEGSPLSLGQ-----SQHDLQSASFSLWSLQTSQSHSSLSLSSVLTDPSSPG 1268
Db 582 --PRQAAPQMLQQOQPRLLISVQTMQRGNMNCAGF-----QAHQMLRAQNA--RIPG 629
QY 1269 KLDPAPOPPEPEPEDEAESPDLOAFWFNISAOQPCNAATPPPLAVSEDOPTSPQOLA 1328
Db 630 -----IPRHSGFOYSMMQPHLQROHNS-----PGHAGFPFVVSVHNTTINPTPTTA 676
```


Db 268 LNSASILPEMEGLS-----EFTEYLSE-SVEVPSPFDILE 301
QY 280 PEPVSGG-----LYGDDTEL-----GAEDX-----LPLXDS 307
Db 302 P-PTSGGLKSKPCYIFPGGRGDSALFAVNGFNMLNGSGERKSCFWKLRHLDRVDS 360
QY 308 PVISAL---DCPSLANATAFSLADDSTSTISFASPTSPVPGVSGVLQDSFDLNNGSD 364
Db 361 ILLTHIGDNDLPGIN-----SMLQRTIAEL-----E 386
QY 365 ABEEMETQSSDF-----PPLSTOPAP-----DOSSTIQ-----LH 395
Db 387 BEQSGSTNSDMKNLSPDLGVVFLNVPENLKNPEINIKMRKRIEACFLQVNLKLS 446
QY 396 PATSP---AVSTTSPAV-----SLVSPA-ASPEISPEVCPAASTVSPAVFS 440
Db 447 MKPELFRSNGTIDPVLFOKMGVGLMYVLPNPKVSKEMQYFMOQWTGNTKDKAEF- 505
QY 441 VVSPASSAVLP---AVSLEVPLTASVTSKASPVTSPPAAFTASPAKNDVSSFLETTAD 497
Db 506 -----ILPNGQEVDFPISYLTSSV-----LIVMHPANPAEKIIRVLPNGNST 548
QY 498 VEEITGEGLTASGSDVMRRIRATPEEVRLLPLQHGWRREVRIKSGHRWOGETWYGPCG 557
Db 549 QYNIL-EGLEKULHDLFKQPLATQKDLTGQVTPVVKQLKQADRS-ES-----LKPAA 603
QY 558 KMKOFPEVIKYLNRNLVHSVRREHFSRPRMPVGDFFEEEDTPEGLQWVLSABEIPSR 617
Db 604 KPLPS-----KSVRKE-----SKEETPE-----VTKNHVKEPK 633
QY 618 IQA-----ITGKRGRPNTE---KAKYKVPKVRGRGRPKVKITELLNKTNNRPLKLEA 671
Db 634 VESKEKVMVKDKPKVKTETKPSVTEKEVP--SKEEPSVKAEVAE-KQATDVKP--KAAK 688
QY 672 OETLNEDKAK---TAKSKKWKQVORGECLTTTQGOARNKQOETKSLKH--KEAKK 726
Db 689 EKTVKETKVPEDKKEKEKPKKVAKKEDTPIKKEKPKKEVKKEVKKEIKKEEK 748
QY 727 SXAEKEKGTQOEKLKEKVRKEKVKMKEKE---EVTK-----AKPACKADKTLA 775
Db 749 EPKEVKETTPKEVKKEKKEKKEKKEKKEKKEKKEKKEKKEKKEKKEKKEKKEKKEK 808
QY 776 TORR---EERQKQMILEMKKPTEDMCLTDHQPDPFSRVPGLTLPSPAFSDCLTIVE 832
Db 809 LKPKVPKEESVKDSV--AAGKPKKGIKVIKKEGAAEAAVAVGTGATTAAMAAA 866
QY 833 FLHSGKVLGPDPAKDVPSLVQGLCOGDSLOEVQDILLVRLKLAALHDPGPPSYCQS 892
Db 867 GIAATG-----PAKELEA-----ERSLMSSPEDLTQDFEEL----- 897
QY 893 LKILGEKVEIPLTRDNVSEILRCFLMAYGVXPAL-----CDRLR-TOPFOAQPPQKAA 946
Db 898 -----KAEVDVTKD-----TKPQLELIEDEKLEKETEPEVAYVYQKERE 937
QY 947 VLAFFVH---ELNGSTLIINEIDKTLGSMSSYRNKNWIVEGRRLKTLVLAKRGRSRVE 1003
Db 938 VTKGPAESPDEGITTTEGECEQTPPELEPEKQ-----GVDDIE 978
QY 1004 MGRPEECLGRRSSRIMETSMEHEEE-----ERSIAAVPGRGRGRDGE--VDATASSIP 1057
Db 979 -----KPEDEGAGFESSETGDYEEKAETEAEPEPEDEGEHVCVSASKHS 1024
QY 1058 ELERQIEKLSKQLFRFKLLHSSQMLRAVSLGQDRYRRYVWVLYLAGIFVEGTEGNLV 1117
Db 1025 PTEDESAAKAEADAVIREK-----RESVAGGDDRAED-----MDEAIKEGAEARQS 1070
QY 1118 PEEVIKKTETSLKVAHAASLNP-----ALFSMKMELAGSN-----TT-----A 1155
Db 1071 BEEA---DEEDKAEDAREEYEPKMEAEADYMAVVDKAAEAGGAEQYGLTTPTKQLGA 1128
QY 1156 SSPAR--ARSRLTKTKPFQMQRPHFKSPVRGODSEQOQAQLOPE-----AQLHVP 1203
Db 1129 QSPGREPASSIHDETLPG-----GSESEATASEENREDQPEEFTATSGYTQSTIEIS 1181

QY 1204 AQOPQLOLQOQSHKFLQEGSGPLSL---GOSQHDLSQSAFLSWLSQTSQSHSLSSSVL 1261
Db 1182 SEPTP-----MDENSTPRDMSDETNEETESPQEFYNTIKYESSLYSQEYS 1229
QY 1262 TP-----DSSPGK---LDPAPSQPPEPEPEDEAESPDQLQAQFWFNISQOM 1303
Db 1230 KPADVTPUNGSESGKTDATDGKYNASASTISPPSSMEEDKFRSALRDAYCSEVKAST 1289
QY 1304 PCNAAPTPLAVSEDQPTSPQOLASSKPMNRPSAANPCS--PVQFSSTP----- 1351
Db 1290 TLDIKDSTF-AVSEKVSPSKPSLSPPSPS-PLEKTPLGERSVNFSLTPNEIKVSABAE 1347
QY 1352 LAGLAP-----KRRAGDPGEMP--OSPTGLGQPKRR----- 1380
Db 1348 VAPVSEPTQBVBEHCASPEDKTLVVVSPSQSVTGSAGHTPYQSPST---DEKSHLPT 1404
QY 1381 ---GRPPS-----KFFKQMEQRYLTQLTAQVPVPPCMCGWWMIIDPEMLDAMLKALHPRG 1432
Db 1405 EVIEKPPAVPVSFBSFSDAKDENERASVPMDEP-----VPDSE--SPIEKVLS- 1451
QY 1433 IREKALHKLHKNRDFL---QEVCLRPSADPFE-----PRQLPAFQE--GIMSWSPK 1480
Db 1452 LRSPLIGSEAYSESFLSADDKASGRGAESPFEESKQSGQSDQVSPVSEMTSTSLYQDK 1511
QY 1481 EKYETDIAVLQWVEELEQRVIMSDLQIRGWTCPSPDSTREDLAYCEH-----LSDSQEDI 1536
Db 1512 QEGKSTDFAPIK--EDFGQEKTTDVEAM-----SSQPALALDERKLGDSVSTQIDV 1561
QY 1537 TWRGPREGL-----APQKTYNPLDLAVMLRLAALQNVKRRVLRPLWPTHVLEK 1589
Db 1562 SQGSEKEDTKMSISEGTVSKSATPVDEGVAE-----DTSGHMEGV 1603
QY 1590 ALLSTPNGA---PEGTTTETS 1607
Db 1604 ASVTASVATSSPPEPTDDVS 1625

RESULT 35

US-09-418-710-62
; Sequence 62, Application US/09418710
; Patent No. 6596482
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042001
; CURRENT APPLICATION NUMBER: US/09/418,710
; CURRENT FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: Fast-Seq for Windows Version 4.0
; SEQ ID NO 62
; LENGTH: 45
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-418-710-62

Query Match 2.8%; Score 273; DB 4; Length 45;
Best Local Similarity 97.8%; Pred. No. 2.6e-14;
Matches 44; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1652 CLVCRKGNDBFLLLCDGCRGCHYCHRPKMEAVPEGDMFCTVC 1696
Db 1 CLVCRKGNDBFLLLCDGCRGCHYCHRPKMEAVPEGDMFCTVC 45

RESULT 36

US-09-839-479-61

```
; Sequence 61, Application US/09839479
; Patent No. 6727222
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042002
; CURRENT APPLICATION NUMBER: US/09/839,479
; CURRENT FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 61
; LENGTH: 45
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-839-479-61

Query Match      2.8%; Score 273; DB 4; Length 45;
Best Local Similarity 97.8%; Pred. No. 2.6e-14;
Matches 44; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1652 CLVCRKGDNDFFLLCGDCRGCHYCHRPKMEAVPEGDWFCCTVC 1696
Db 1 CLVCRKGDNDFFLLCGDCRGCHYCHRPKMEAVPEGDWFCCTVC 45

RESULT 37
US-09-538-092-1262
; Sequence 1262, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
; APPLICANT: Giot, Loic
; APPLICANT: Mansfield, Traci A.
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538,092
; CURRENT FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CuratPatSeqFormatter Version 0.9
; SEQ ID NO 1262
; LENGTH: 3969
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: Polypeptide Accession Number Q03164
US-09-538-092-1262

Query Match      2.8%; Score 272; DB 4; Length 3969;
Best Local Similarity 17.7%; Pred. No. 5.1e-11;
Matches 395; Conservative 250; Mismatches 706; Indels 880; Gaps 98;

Qy 17 LPPAP-----AASGLKPS-
Db 40 LPPGPPVGGGPGAPPSPPAVAAAAAAGSGAGV-----PGGA 78

Qy 62 STVSHTTTSGTLNAPHSSTSH---LHPSVAVDCLWNYSQYPSANPGSNLKDPPLLSQ 118
Db 79 AAASAASSSSASSSSSSSSASSGPFALLRVGFGDA-----ALQVSAATGTLNRRFVFG 134

Qy 119 FSGQYPLNGILGGS-----RQPS-SPSHNTN-----LRAGSQKFW 153
```

```
Db 135 ESGGG-----GGSGDEQFLGFGSDEEVRVRSPTSPSVKTSRPRGRPRSGSDRNS 187
Qy 154 A-----NGTHSPMGLN-----FDSQEL-----YDSPDQ 177
Db 188 AILSDPSVSVPLNKSETKSGDKIKKDSKSIKKRGRPPPTPGVKIKITHGDKISELPGK 247
Qy 178 NFEVCSGIHPDEAAEKEMTSVVAENGTGLVCSL-----ELEEXOPELKMCVNGSVVP 230
Db 248 NKEDSLKKIKRTPSATFQQATKIKKLKAGKLSPLSKFKTKGLQIGRKGVIIVRRGRPP 307
Qy 231 SVESLHGEVSVLPDPTVSCLDPSHLPDQLEDTPILSEDS-----LEFPN----- 276
Db 308 STERIKTPSGLLIN---SELEKPKVKRDKEGTPTLKEDKTVVRQSPRIKPVRIIPS 363
Qy 277 -----SLAPEVSGGLYGI-----DDTELMAEDKLPLXD-----SPVTSALDCPSLN 319
Db 364 SKRTDATTIAKQLLQRAKGAQKIEKAAQLOGRKVKTVQVKNIRFOFIMPVSAISRIIK 423
Qy 320 NATAFSLADD-----SOTSTSIASPTSPVPLGESVLQDNSFDLNGSD-AEQ 367
Db 424 TPRPF--IEDEDYDPIKIALESTPNRSFAPSACGSEKSSAASQHSQSSDSSRSS 481
Qy 368 EEMETQSDPPPSLTQAPDQ--SSTIQHPAATSPAVSP----- 404
Db 482 PSVDTSDSQASEBIEQVLPEERSDTPVHPPLPISQSPENESNDRRRRYSVSERSFGSR 541
Qy 405 TTSPAVSLVSVPAASPEISPEVCPAASTVSVPAFVSVPASSAVLPAVSLVPLTASVT 464
Db 542 TTKKLSTLQSAPOQOOTSSTSP--PPLLTTPPPP-----LQPASS-----ISDHTWLMPEPT 589
Qy 465 SPKASPVTSAAAPPTASPAKNDVSSLETT-----ADVEITCEGLTAS 509
Db 590 IPLASPF-LPAS-----TAPMQGRKXSIILREPTFRWTSLKHSRSEPOYSSAKYAKEGLRK 645
Qy 510 GSGDWRRRIATPEEV-----RL--PLQHGWRREVRKKGSHRWOGETWY 552
Db 646 PIPDNFRPPPLTPEDVGFASGFSASGTAAGARLPSLHSGTRFDWM----- 691
Qy 553 YGPGCKGMKOPPEVIVKYLNRNLVHSVREHFSFSPRMPVGDFFERDTPEGLQWVLSAE 612
Db 692 -----KRSP-----LRAPR-----FTPSBAHSRIFESVTLPSNRTSAGTSSS 729
Qy 613 EIPSRIOAITKGRG---PRNTEKAKTKEVPKVRGR-----GRPKVK----- 653
Db 730 GVSNR-----KRKKVFSPIRSEPRSPSHSMRTFSGLSSELSPLTTPPSVSSLSIS 783
Qy 654 ----ITELANKDNRLPKLEAQETLNEEDKAKIAKSKXMR----- 691
Db 784 VSPLATSALNPTFTFPGHSL----TQSGESAENQRPRKQTSAPAEPPSSSPTLPFWF 839
Qy 692 ---QKVORGECLTTIQGAARNKRQETKSLKHKEAKKSAEKEKGTQKQELKEKVKRE 748
Db 840 TPGSQTERG-----RNKKKAPBELSKORDADK--SVEKDKSRERDRERKENKRE 887
Qy 749 KKEVKMKKEK-----EVTKAKPACAKDKTLATORRLBEROKQOQMILEMKKPT 797
Db 888 SRKEKRKKGSEIQSSSALYPVGRVSKKVGEDVATSSAKKATGRKSS----- 937
Qy 798 EDMCLTDHQPDPFSRVFGLTLPFGAFSDCLTI--VEFLHSEFK-----VLGDDPKADVPSL 852
Db 938 -----SHDSGDITSVTLGDDTAVTKILIKKGRGNLEKTNLDLGTAPSL 983
Qy 853 GVLOEGLLCOGDSLGEVQDILLVRLKALHDPGFPYCSQSLKILGEKVSE---IPLTRDN 909
Db 984 E--KEKTL-----LSTPSSSTVKHSTSSIGSLMAQADKLPMTDKR 1022
Qy 910 VSEILRCFLMAYGVKPALCDRLRTPFQAQPPQKAAVLAPFVHELNGSTLIINEIDKTL 969
Db 1023 VASLTK-----KAK 1031
Qy 970 ESMSSYRKNKWI VEGRLRLUKTVLAKRTGRSEVEMGRPECLGRRRRSRIMEETSGMEE 1029
```



```
Db 1384 -----NGSSQKQI--PAD-GVHRIRV---DFKEDCEAENVWEMGGLGI---LTSVP--- 1426
Qy 1599 PEGITTEISYEITPRIRWQTLQCRSAHV-----C-----LCIGHLSIAWEK 1645
Db 1427 -----ITPRVVCF-----LCASSGHVEFYVCQCCEPFHKFLEENERPL--ED 1468
Qy 1646 SVNKVTCLVC-----RKGDNDEFLLLCDGDRCHLYCHRPKWEAVP---EGDWFCIV 1695
Db 1469 QLENWCCRRCFKCHVCGHQHATKOLLECNKCRNSYHPECGLGPNYPTKTKKKVWICTK 1528
Qy 1696 CLAQQVEGEFTQKPGFPKRGOKRSGYGLNFSEGDRRRVLLKGRSPAAGPRYSERL 1755
Db 1529 CVRCKSCGSTT-----PGKGWDAQWHDPSLCHDCA---KLFAKNFCLCDKCYDDDDY 1580
Qy 1756 SPSKRRRLUM--RNHSLDTTCETILMEM-----ESHDAW----- 1789
Db 1581 E-SKMOCGKCDRWVHSK---CENLSDMEYILSNLPESVAYTCVNCCTERHPAEWRLALE 1636
Qy 1790 -----PFKE-----PVNPRLVSG 1802
Db 1637 KELQISLKQVITALLNSTTSHLLRYROAAKPPDLNLPETESIPSRSPGDPDPVLTEV 1696
Qy 1803 YRRIKPNMDFSTWRERLLRGYTSSEFAADALLVFDNCOTFNEDDSEVGKAGHIMRRF 1862
Db 1697 SKQDDQQLDLLEGVKRKMVDQNYTSVLFSDDIVKIIQAAINSDDGQOPEIKKANSWVKSF 1756
Qy 1863 FESRWEEFY 1871
Db 1757 FIQWERVEF 1765

RESULT 40
US-09-513-999C-7578
; Sequence 7578, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert, A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; Patent No. 6783961
; FILE REFERENCE: 59.US2.REG
; CURRENT APPLICATION NUMBER: US/09/513,999C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
; SEQ ID NO 7578
; LENGTH: 52
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-513-999C-7578
```

```
Query Match 2.6%; Score 256; DB 4; Length 52;
Best Local Similarity 100.0%; Pred. No. 7.2e-13;
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 370 METQSSDFPPLSQPAPQDSSTIQLHPATSPAVSPPTSPAVSLVVSPPAASP 420
Db 1 METQSSDFPPLSQPAPQDSSTIQLHPATSPAVSPPTSPAVSLVVSPPAASP 51
```

```
RESULT 41
US-09-264-512B-2
; Sequence 2, Application US/09264512B
; Patent No. 6610508
; GENERAL INFORMATION:
; APPLICANT: Hentze, Matthias W.
; APPLICANT: De Gregorio, Ennio
; TITLE OF INVENTION: TRANSLATION DRIVER SYSTEM AND METHODS FOR USE THEREOF
; FILE REFERENCE: 9882-004
```

```
; CURRENT APPLICATION NUMBER: US/09/264,512B
; CURRENT FILING DATE: 1999-03-08
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 1560
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-264-512B-2

Query Match 2.6%; Score 252; DB 4; Length 1560;
Best Local Similarity 19.2%; Pred. No. 4.1e-10;
Matches 346; Conservative 226; Mismatches 650; Indels 576; Gaps 86;

Qy 18 PPAPAAAGSLKPS-----PSSGEGLYTNGSP-MNFPQOQKSLNGDVNVNGLSTVSHTTTSGI 72
Db 18 PPSAAARVQSAAPARPGPAAHVYPAGSQVNMIPSQ-----ISYPASOGA 62
Qy 73 LNSAPHSSSTSHLHPHVAYDCLWNYSQYPSANPGSNLKDPLLQSFSGGOYPLNGILGG 132
Db 63 YYIFGQGRSTYV--PTQOYPV---QPCAGCFYFGAS---PTEFGTYAGAYYPAQGV--- 111
Qy 133 SRQSPSSHNNTNLAGSQKFWANGTHSPMGLNFDSQ-----ELYDSFFPON--- 178
Db 112 -----QQFTGVAPAPVLMNQPPQIAPKRERTIRID--PNQGGKDI 152
Qy 179 FEEVCSGIHPDEAAAEKEMTSVVAENGTLVCSLEEXQPELKMCGVNGSVPSVESL--- 235
Db 153 TEEIMSG-----ARTASTPTPQTGGGL-----EPQA-----NGETPQAVIVRP 192
Qy 236 --HOEVSVLVPDPTVCLDDPSHLPDOLE-----DTPILSEDSLEPFNSLAPEVPS 284
Db 193 DRSQGAIIADRFQ---LPGHEHSPSESQSPSSPPTSPSPVL-EPGSEP--NLAVLSIP 246
Qy 285 GGL-----YGIDDTLMGAEDKLPXDSPVISALDCPSLNNATAFSLADDSOTSTIFA 339
Db 247 GDTWTTIQMSVEESTPISRETGEYRLSPBPTPLAEPILEVEVTLKSPVPESEFSSSPLQ 306
Qy 340 SPTSPPVLGSLVDNDSFDLNGSDAEQEMETOSSDFPSPSLTOPAPDOSSTIQLHPATS 399
Db 307 APT--PLASHTV---EIHBPNGMVPSEDLPEVSS--PELAPPPACPSESVPPIAPTAQ 359
Qy 400 P-----AVSPPTSPAVSLVVSPPAASPE-----ISPVECPAAASTVSPAVSFVVSP 444
Db 360 PEELLNGAPSP--PAVDL--SPVSEPEQAKEVTASVAPTIFSATPATAP---SATSP 411
Qy 445 A-----QPPEVIKYLNRNLVHVSRRHFSPRMPVGDFFFEERDTPEGLQWVQLSAEIP 477
Db 412 AQEEMEEEBEEREEAGEAGEAESEKGBELLPESTPIPANLS-QNLEAAAAATQAVS 470
Qy 478 PPTASPANKQVS-----SFLETTADVEEITGE-----GLTASGSDVMRRRIA 520
Db 471 VPKERRIKELNKKEAVGDLIDAFKEANPAVPEVENQPPAGSNPGPSESGSVGPPR---- 526
Qy 521 TPEEVRLPLQHW-----RREVRIKGGSHRWGETWYVGPCCGRMK----- 561
Db 527 -PEEA-----DETWSKEDKIHNAENIQPGEOKYKSDQWKPNNLEKKRYDRFLGQ 581
Qy 562 -----QPPEVIKYLNRNLVHVSRRHFSPRMPVGDFFFEERDTPEGLQWVQLSAEIP 615
Db 582 FIFASMCKPEGLPHISDVLDKANK---TFLRPL-----DPTRLQINGCPDPTF 628
Qy 616 SRI---QAITGKRGRPRNTEKAKTEVKVVRG-----RGPPPKVKITEL-----L 658
Db 629 SFANLGRITTLSTRGPPRGPGGELPRGPOAGLPRRSQQGPRKEPRKIIATVLMTEITKL 688
Qy 659 NKTEN--RP-LKKLEAOETLNEED-----KAKIAKSKKXKROKQVQSGECLTT 702
Db 689 NKAENKWPSSKRTAADKDGEEDADGSKTQDLFRVRVRSILNKLTTPQFMQLMKQVTLA 748
Qy 703 IQGOARNKR-----KQETKSLKHKE-----AKKKSXAEBKKGKTKOEKLKEKVK 746
Db 749 IDTEERLKGVIDLIFEKAISEPNFSAVANNMCRCLMALKVPTTEKPTVTNVFRKL--LLN 806
```



```
Db 1978 FEHKSAAK---KQPOEKGKVRVEKEKGPILTQREAAKTEN-----QTIKR 2019
Qy 697 BECLTTTIOGARNRKQETKSLKHKEAKKSAEKEKTKQEKLEKVKREKKEKVKWK 756
Db 2020 GORL-PTVTGAEKRGVRVSS1-----GVKEDAAGGKEKVLSHK1PEPVQSVPEE--ESH 2072
Qy 757 EKEEYTKAPACKADKTILATORRLEERQKQOMILEMKKPTEDMCLTDHQPLPDPFSRVP 816
Db 2073 RESEVPKEKWA-----DEQGMDLQISPRKTSITDFSEVIKQBELENDKVKQ 2119
Qy 817 LTLPGASDCDLTIVEFIHLS-FGKVLGPDPAKD--VPSLGLVQEGLLCGDLSGEVODLL 873
Db 2120 FRLSETEKAQLHLQDLVTSPTNFTFPLDYMKDEFLPALS-LQSGAL-DGSS-----2169
Qy 874 VRLKKAALHDPGFPPYCSOLKTLGKVEIPLTRDNVSEILRCFLMAYGVAPALCDRLRT 933
Db 2170 -----BSLK--NEGVAGSP-----CGSLMEGTPTQISEESYK 2199
Qy 934 QPFOAQPPQKAAVLAFPVHELNGSTLIINEIDKT-----LESMSYRKNKWIVEGRURR 988
Db 2200 HEGLAETPTSPESLSFPKSEEGTGETKSTKTETTEIRSEKEHPTTKDITGGSEER 2259
Qy 989 LKTV-----LAKRTGRSEVEMGRP-----EECLGRRRRSRIMEETSGMEE--BEE 1032
Db 2260 GATVTEDSETSTESFOKEATLGS PKDTSPKQDDCTGSCSVALAKETPTGLTEAAACDEG 2319
Qy 1033 ESIAAVPGRGRDGEV-DATASS-----IPELEROIEKLSKQLFRRKLLHSSQML 1084
Db 2320 QRTFGSSAHKTQTDSEAEQESTATSDTKALPLPEASVKTDGTGTSK---PGOVIRSPOGL 2376
Qy 1085 RAVSLGQDQRYRRYVLPYLAGIFVETGEGNLPVEVIKK-ETDSLKVAHAASL--NPAL 1141
Db 2377 ELALPSRD-----SEVLSAVADDSLAVSHKDSLEASVPL 2410
Qy 1142 FSKMELAGSNNTASSPARAKSRPLTKPGFMQPHFKSPVRGQDSEQFQAQLOPEAQL- 1200
Db 2411 -----EDNSSHKTPDSLPSPLK-----ESPCRDLSLESSP---VEPKMKAG 2448
Qy 1201 -----HVP---AQPPQLOLQLOSHKGFL--FQEGSPLSLGOSQHDLSOSAFLSMISQTS 1251
Db 2449 IFPSHFPLPAVAKTELLTEVASVRSRLRPDPGSAEDDSLEQTSLMESSGKSPLSPTDP 2508
Qy 1252 HSSLSSSVLPDSSPGKLDPAPOPPPEEPDEAES-----SPDLQAFWNISAQM-- 1303
Db 2509 SSEEVSVEV-TPKTTDVS-TPKPAVIEHCAEDDSENGINEKRRFTPEEEMFKMTKIRKFD 2566
Qy 1304 -----PCNAAPTTPPLAVSEDDP--TPSPQOLASSKPMNRPSAAN 1340
Db 2567 ELEQEAQKRDYKKEPKQEESSSSDDPDADCSVDVDEPKHTGSGED-ESGVPLVLTSESR 2625
Qy 1341 PCSPVQFSGSTPLAGLAPKRRAGDPOEMPOQSPTGLGQPKRRGRPPSKPKQMEQRYLTOLT 1400
Db 2626 KVSSSESEPELAQL---KKGADSGLLP-----EPVIRVQPPS-----2660
Qy 1401 AQVPPEPCSGWMTIPDPEMLDAMLKALHPRGIREKALHKLHKNHRDPLQEVCLRPSADP 1460
Db 2661 ---PLPSSMDSN-----SSPE-----EVQFPVVSKQYTP-----KMNEDT 2693
Qy 1461 IFEPRLPAFOGIMSWSPKPKTYETDLAVLQWVEELEQRVIMSDLQIRGTCPSPOSTR 1520
Db 2694 QEEPGK-----SEEEKDSHSLA-----EDRAVVS-----TEA 2721
Qy 1521 EDLAYCHELSQEDITWRGPGREGLAPOKRTYUPLDLAVMRLAALAEONVRRYLRLEPLW 1580
Db 2722 EDRSYDKLNRDTPQKICDGHGCEAMSPS--SSARPVSSGLQS-----2762
Qy 1581 PTHEVWLEK-----ALLSTPNCAPGEGTTEISYEITPRIRIWRQTLQRCRGAHVCL 1632
Db 2763 PTGDDVDVQPIVYKESLALQTHEKDTGELDVSRAESPADCFSEFSSSSLPHCLV 2822
Qy 1633 CLG-HLERSIAWEKSVNK 1649
Db 2823 SEGKELDEDISSTSIQK 2840
```

RESULT 43

```
US-09-949-016-7659
; Sequence 7659, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: C1001307
; CURRENT FILING DATE: 2000-04-14
; PRIOR FILING DATE: 2000-10-20
; PRIOR FILING DATE: 2000-10-20
; PRIOR FILING DATE: 2000-10-03
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7659
; LENGTH: 2753
; TYPE: PRN
; ORGANISM: Human
US-09-949-016-7659
```

```
Query Match 2.4%; Score 240; DB 4; Length 2753;
Best Local Similarity 20.6%; Pred. No. 9.2e-09;
Matches 315; Conservative 173; Mismatches 571; Indels 472; Gaps 70;
```

```
Qy 142 NTNLRAGS---QKFWANGTHSPMGLNFDLSQELYSDFPDQNFEEV-----CSGI 186
Db 1357 NEILRSTGCTRDESSVQSSRSERGL---VEEWWIVSDEEIEEARQAKAPLEITEYPCVEV 1413
Qy 187 HPDEAAEAKEMTSVVAENGTLGV-----CSLELEEXQPELKMCGYNGSVPSV-ESL 235
Db 1414 RID-----KEIKGVKEKSTGLVNYLTDDLNTCVLPKEQLQTVQDKAGKCEALAVGRSS 1469
Qy 236 HQEVSVLVPDPTVSCLDPDGHLPDQLEDPTI-----LS 268
Db 1470 EKEGKDIPDETQS--TQKHKPSLGIKKPVRRKLKEQKQKERGLQASAEKAEKKGSS 1527
Qy 269 EDSLPPNSLAPEVPVSGGLYGDIDTELGAEDKPL--XDSPVLSALDCPL-NNATAPS 325
Db 1528 EESLGEDPGLAPEP-----LPTVKATSPLEETPIGSIKDKVKALQ 1568
Qy 326 LLADDSOTSTISIPASPTSPVLGSEVLQDNSFDLNGSDABQEBMETQSSDFPPLSTOPA 385
Db 1569 KRVEDEOK-----GRSKL-----PIRVKGEDVPKKTTHR 1598
Qy 386 PDQSSSTIQLHPATSPAV-----SPTTSPAVSLVSPAPASPEISPEVCPAAST---V 433
Db 1599 P-----HPAASPSLKSERHAPSPKPTERHSTLSSSAKTERHPPVSPSSKTEKHP 1650
Qy 434 VSP-AVFSVSVSPASSAVLPAVSLVPLTASVTSKASPVTS-PAAAPTASPAKDVSSF 491
Db 1651 VSPSAKTERHSPASS--SKTEKHSVPSPSTKTERHSPVSTKTERHPPVSPSPGK-----1703
Qy 492 LETTADVEEITGEGLTASGCDVMVRRRIATP--EEVRLPLQHGWRREVRIRKKGSHRWQGE 549
Db 1704 -----TDRKPPVSPSGORTEKHPVPSPGRTEKRLPVSPSGRTD---KHQPVSTAGK 1750
Qy 550 TWYY---GPCGKRMKQFPVVIKYLRSRLNVHSVRREHFSFSPRMPVGVDFFEERDTPEGLQW 606
Db 1751 TEKHLVPSPSGKTEKQPP-----VSPTSKT-EBIES-----1780
Qy 607 VQLSABEIPRIQAITKGRGRPRNTEKAKTEKVPKVRGRGRPPKVKITELLNKTDRPL 666
Db 1781 -TMSVRELMKAFQS-----GQDPSKHKTLGLFEHKSAAQ---KQPOEKGKVRVEKEKGPIL 1831
Qy 667 KCLEAQETLNEEDKAKIAKSKKMKRQKVORGECLTTTQGOARNKRKQETSLKHKEAKKK 726
```

Db 1832 TQREAOAKTEN-----QTIKRGQRL-PVTGTABSKRGVRVSSI-----GVKK 1871
QY 727 SXAEKEKGTOKKLEKVKREKKEKVKMEKEEVTAKPACKADKTATQRLERQKQ 786
Db 1872 EDAAGGKVKLHKIPVQSVPEE--ESHRESEVPKEMA-----DEQGM 1916
QY 787 QMILEMKKPTEDMCLTDHQPDPDFSRVGLTIPSGAFSDCLTIVEFLHS-FGKVLGDP 845
Db 1917 DQISPDRTSTDFSEVVKQELDNKYQOFRLESETEKAQHLHQVLTSPTTFPLDY 1976
QY 846 AKD--VPSLGVLEGLLQCGDSLGEVQDLLVRLKALHDPGPPSYCQSLKILGEKVSBI 903
Db 1977 MKDEFLPALSLQSGAL-DGSS-----ESLK--NEGVAOS 2007
QY 904 PLTRDNVSEILCFMAYGVXPALCDRLRTQFPAQPOOKAAVLAPVHNLGSLIIN 963
Db 2008 P-----CGSLMEGTPOITSEESYKHEGLAETPETSPELSFSPKKEEQTGENK 2056
QY 964 EIDKT-----LESMSYRNKNIIVEGLRLRLKTV-----LAKRTGSEVEMGRP-----1007
Db 2057 ESTKTETTEIRSEKHPHTTKDITGSEERGATVTEDSSTESTESFQKEATLGSFKDTSPK 2116
QY 1008 --EECLGRRRRSRIMETSGMBEE---EEESIAAVPGRRRRGRDGV-DATASS-----1055
Db 2117 RODDCTGSCSVLAKETPTGLTEEAACDEGRTFGSSAHKTQDSEVQESTATSDTKAL 2176
QY 1056 -IPELERQIEKLSKRLQFPRKXLLHSSQMLRAVSLGQDRYRRYVWLYPLAGIFVEGTG 1114
Db 2177 PLPEASVKTDGTESK--POGVIRSPQGLELALPSRD-----2211
QY 1115 NLVPEVVIK-ETDSLKVAAHASL--NPALFNMKMLAGSNNTASSPARARSRPKTRPG 1171
Db 2212 ----SEVLAVADDLSLAVSHKDSLEASPVL-----EDNSHKTPTDSLPSPLK----2255
QY 1172 FMQPRHFKSPVRGQDSEQQAOLQPEAQL-----HVP-----AQPQLOLQLOSHKGF--1221
Db 2256 ----ESPCRDLSESP---VEPKMKAGIFPSHFPLPAVAKTELLTEVASVRSRLR 2305
QY 1222 EOGSPSLGQSHDLSQAFSLWSLQTSQSHSLSSSVLTTPDSSPGKLDPAPOPPBP 1281
Db 2306 DPGSAEDDSLQTSIMLESSGKSPSPDPSPSEESVSEYV-TPKTDVS-TPKPAVHECA 2363
QY 1282 EPDEAES-----SPDLQAFWNISAOQ-----PCNAAPTP 1312
Db 2364 EEDDSGKKGKFTPEEMFKMTIKMFDELEQEAQKQKDYKBPQKQESSSSDDPAD 2423
QY 1313 LAVSEDP--TPSPQOLASSKPMWRPSAANPCSPVQFSFTPLAGLAPKRACDPMQPS 1370
Db 2424 CSVDDEPKHTGSD-ESGVPLVTSESRKVSSSESEPELAQL---KKGADSGLLP--2477
QY 1371 PTGLGQPKRGRPPPKFFKQMEQRYLTQLTAQVPPMCGWWITPDPEMLDMLKALHP 1430
Db 2478 ----EPVIRVQPPS-----PLPSMDSN---SSPE-----2500
QY 1431 ROIRKALHKLHKLHDFLOEVCLRPSADPIEPRLQPAFQGINWSWSPKETYETDLAV 1490
Db 2501 ----EVQFQPVVSKQVTF-----KXNEDTQEPGK-----SEEEKDSHSLAE 2539
QY 1491 LQWVELEQRVMSDLQIRGWTCPSPDSTRE 1521
Db 2540 DQARELD----FTBEQIHQIRIENPNSLQD 2566

RESULT 44

US-09-949-016-7660
; Sequence 7660, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016

; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 7660
; LENGTH: 2753
; TYPE: PRT
; ORGANISM: Human
; US-09-949-016-7660

Query Match 2.4%; Score 240; DB 4; Length 2753;

Best Local Similarity 20.6%; Pred. No. 9.2e-09;

Matches 315; Conservative 173; Mismatches 571; Indels 472; Gaps 70;

QY 142 NTNLRAGS---QKFWANGTHSPMGLNFDSPQELYDSPDQNFEEV-----CSGI 186
Db 1357 NEILRSCTTRDESSVQSSRSGRL---VEEWVIVSDEEIEEARQKAPLEITEYPCVEV 1413
QY 187 HPDEAAEKEMTSVVAENGTLV-----CSLEEXQPELXKMGVNGVSPV-ESL 235
Db 1414 RID-----KEIKGKVKDSTGLVNYLTDLNTCVLPKEQLQTVODKAGKCEALAVGRSS 1469
QY 236 HOEVSVLPDPTVSCLDPPSHLPQLEDPTI-----LS 268
Db 1470 EKEGDIIPDETQS--TKQHKPSLIGIKKPVVRKUKEKQKQKEGLQASAEAKLKGGS 1527
QY 269 EDSLFPNLSLAPEVPVSGGLYGIDDTLMGAEDKFL--XDSPIVIALDCPSL-NNATAPS 325
Db 1528 EESLGEDPLAEP-----LPTVKATSPLEETPIGSIKDKVKALQ 1568
QY 326 LLADDSQTSIFASPTSPVPGESVLDQNSFDLNGSDAEQEEEMETQSSDPPSLTQPA 385
Db 1569 KRVEDEQK-----GRSKL-----PIRVKGEKEDVPKKTTHR 1598
QY 386 PQDSSTIQLHPATSPAV-----SPITSPAVSLVSPAPASPEISPEVCPAAST---V 433
Db 1599 P-----HPAASPLSKERHAPGSPKTERHSTLSSAKTERHPPVSPSKTEKHSP 1650
QY 434 VSP-AVFSVSPASSAVLPAVLEVLTAASVTSPKASPVTS--PAAAFPTASPANKDVSSF 491
Db 1651 VSPSAKTERHSPASS--SKTEKHSVPSPSTKTERHSPVSSSTKTERHPPVSPSGK----1703
QY 492 LETTADVEEITGEGLTASGSDVMRRRIATP--EEVRLPQHGWRREVRKKGSHRWQGE 549
Db 1704 -----TDKRPVSPSGRTEKHPPVSPGRTEKRLPVSPSGRTD---KHQPVSTAGK 1750
QY 550 TWYI---GPCGKMKQFPEVILKLSRNLVHSVRREHFSPSPMPVGDPEEDRTPEGLQW 606
Db 1751 TEKHLFVSPGKTEKQPP-----VSPTSKT-ERIE-----1780
QY 607 VQLSAEIIPSRIQATIGRGRPRNTEKAKTEKVPKVRGRGRPPKVKITELINKTDNRPL 666
Db 1781 -TMSVRELKAFQS-----GQDPSKHKTGLEHKSQKQ---KQPOEKGVKVRKEKGPIL 1831
QY 667 KKLAEQETLNEDKAKIAKSKKMRQKQVORGECLTTIOGARNRKKQETKSLKHKEAKKK 726
Db 1832 TQREAOAKTEN-----QTIKRGQRL-PVTGTABSKRGVRVSSI-----GVKK 1871
QY 727 SXAEKEKGTQKLEKVKREKKEKVKMEKEEVTAKPACKADKTATQRLERQKQ 786
Db 1872 EDAAGGKVKLHKIPVQSVPEE--ESHRESEVPKEMA-----DEQGM 1916
QY 787 QMILEMKKPTEDMCLTDHQPDPDFSRVGLTIPSGAFSDCLTIVEFLHS-FGKVLGDP 845
Db 1917 DQISPDRTSTDFSEVVKQELDNKYQOFRLESETEKAQHLHQVLTSPTTFPLDY 1976
QY 846 AKD--VPSLGVLEGLLQCGDSLGEVQDLLVRLKALHDPGPPSYCQSLKILGEKVSBI 903

Db 1977 MKDEFLPALS-LQSCAL-DGSS-----ESLK--NEGVAGS 2007
Qy 904 PLTRDNVSEILRCFLMAYGVXVAPALCDRLRTPFOAQPPOQKAAVLAFPPVHELNGSTLIIN 963
Db 2008 P-----CGSLMEGTPOISEESYKHGUAETPETSPESLSFPKXSEBOTGETK 2056
Qy 964 EIDKT-----LESMSYKKNWIVBGRRLRLKTV-----LAKRTGRSEVEMGRP----- 1007
Db 2057 ESTKTETTERSEKHEHTTKDITGGSEERGATVETEDSETSESFOKEATLGSPKOTSPK 2116
Qy 1008 ---ECLGRRRRSRIMEETSGMEE-----EEESIAAVPGRRRRRDEEV-DATASS----- 1055
Db 2117 RQDDCTGSCSVALAKETPTGLTEEAACDQRTFGSSAHKTQDSEVQESTATSDTKAL 2176
Qy 1056 -IPELERQIEKSKQLFRKLLHSSQMLRAVSLGQDRYRRYVWLPYLAGIFVEGTG 1114
Db 2177 PLPEASVKTDTGTESK-----POGVIRSPQGLALPSRD----- 2211
Qy 1115 NLVPEVIKK-ETDSLKVAAHASL--NPALFSMKMELAGSNNTASSPARARSRLPKTKPG 1171
Db 2212 ----SEVLAVADDLSLAVSHKDSLEASPVL-----EDNSSHKTPDSLEPSLK----- 2255
Qy 1172 FMQPRHFKSPVRGQDSEPOQALOPEAQL-----HVP---AQOPQLOLQLOSHKGFL-- 1221
Db 2256 -----ESPCRDLSLESP--VEPKMKAGIFPSPHFPLPAAVAKTELLTEVASVRSLLR 2305
Qy 1222 EQEGSPLSGOSHDLQSALFSLWSQTSOSSLLSSSVLTPDSSPGKLDAPSQPPPEP 1281
Db 2306 PDGSAEDDSLEQTSLEMSGGKSPSPDTPSPSEVSSEYEV-TPKTTDVS-TPKPAVIHECA 2363
Qy 1282 EPDEAES-----SPDLOAFWNISAOM-----PCNAAPTPP 1312
Db 2364 EDDDSSENGEKRFTEEFENFMVTKIKMPDELEQEAQKRDYKKEPKOESESSSDPDAD 2423
Qy 1313 LAVSEDOF--TPSPQOLASSPMRNPSSAANPCSPVQFSSTPLAGLAPKRRAGDPCEMPQS 1370
Db 2424 CSVDDEPKHTGSGED-ESGVPLVTSESRKVSSSESEPELAQL-----KKGADSGLLP-- 2477
Qy 1371 PTGLGQPKRRGRRPPSKFKQMEQRVLTQLTQAQVPPPCSGWWVTPDPEMLDAMLKALHP 1430
Db 2478 ----EPVIRVQPPS-----PLPSSMDSN-----SSPE----- 2500
Qy 1431 RGIREKALHKLHNRDQLQECVLRPSADPIPEPRQLPAFOEGIMSWSPKKTETVETDLAV 1490
Db 2501 ----EVQOPVVSQKYTF-----KWNEDTQEEPK-----SEEEKDSHSLAE 2539
Qy 1491 LQWVEELQRVIMSDLQIRGWTCPSPDSTRE 1521
Db 2540 DLARELD----FTBEEQIHQIRIENPNSLQD 2566

RESULT 45
US-09-949-016-8902
; Sequence 8902, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8902
; LENGTH: 1878
; TYPE: PRT

; ORGANISM: Human
US-09-949-016-8902
Query Match 2.4%; Score 239.5; DB 4; Length 1878;
Best Local Similarity 19.5%; Pred. No. 5.3e-09;
Matches 318; Conservative 208; Mismatches 620; Indels 485; Gaps 74;
Qy 172 DSRPDQNFEEVCSGIHPDEAAAEKMT-----SVVAE-----NGTGLVCSL 211
Db 510 DSFPVSLQAVTDSAMTSLKLEKAMTEPSALIEKSSIQELFEMRVDDKDKIEGVGAATSA 569
Qy 212 BLEEXQPELKMCGYNGSVPSVESLHQEVSVLVPPDTVSCLDPPSHLPDQLEDTPILSBD 271
Db 570 ELDMPFYEDK-----SGMSKYFETSALKEATKSI--EPGSDYVELSDT---RESV 615
Qy 272 LEFNSLAPVPSVGLGIDDTLGMGAEDKLPLXDSVPISALDCPSLNNAATFSLADDS 331
Db 616 HESIDTWSPMHKG-----DKFQTKGESQF---SPQAQAGYSTTLAQSPSPLPEEPS 666
Qy 332 QTSTSIASPTSPVPLGE-----SVLQDNSFDLNNGSDAEQEEMET----- 372
Db 667 SPQBRMTT--IDPKVYGEKRDLSKNKDDLTLSRLSLGGRSAIEQRSMSINLPMSCIDS 724
Qy 373 -----QSSDFPP--SLTPQAPDQSSSTIQLHPATSPAVS-----PTTSPAVSLVVS 415
Db 725 IALGFNFGRGHDLSPLASDILTNTSGSMDEGDDYLPATTPALEKAPCPFPVESKEEQIEK 784
Qy 416 PAASPEISPEV---CPAA-----STVVSFAVSVVSPASS-AVLPAVSLVPLTAS 462
Db 785 VKATGEESTQAEISCESPFKAFKNGTWNAPOLPEMLDLAGTRSLASVADAEEVARR 844
Qy 463 VTSKASPVTSAPAAFTASPAKDV--SSFLB-----TTADVEBITG 503
Db 845 KVSPTSETVEDSRGLPVDENHVIKTSQLEDGLCYCVENKVTVPPLSPVQSENUSG 904
Qy 504 E-GLTASGSGVMRRRIATP---EEVRLPHQHRREVRIKKGSHRMQGETWYVYGPCGR 559
Db 905 ESGTFYEGTDKVRDLATDLSLIEVKLAAAGRVKDFSVDK----EASAHISGDKSGL 959
Qy 560 MKQPEVIK-----YLSRNLVHSVREHPSFSPMPVGDFFPEERTPEGLQWVLSAE 612
Db 960 SKFEDQEKANDRLDTVLEKSEEHADSKHAKTTEE--AGDEIETFGJ--GVTYEQALAK 1015
Qy 613 E--IPSRITQAITKGRPRNTEKAKTKEVPKVGK-----RGRPPKVKITSL-----LN 659
Db 1016 DLSIPTDASSKAKGSLSSVPEIAEVEPSKXVGOGLDPAVQOQ-LDVKISDFGQWASGLN 1074
Qy 660 KTDNRPLK-KLEAQETLNEEDKAK-----IAKSKKMKQKQVQGECLTTIOGAARNK 710
Db 1075 IDRRATELKEATQDMTPSSKAPQEADAPMGVESHMKEGTVKSVTEV-----K 1124
Qy 711 RKQETSKLKHKAKKKXAEKEKTK---OEKLEK-EKVKEKKEKVKMKKEEVTKAK- 765
Db 1125 EKVAKPDLVHQEAVDKEESESGEHESLTWESLKADEGKETSPESLIQDEIAVKLSV 1184
Qy 766 ----PACKADKTATQRLEERQKQMILEMKKPTEDMCLTDHQPDPFSRVPGLTLP 821
Db 1185 EIPCPPAVSEADLAT----DERADVQM--EFIQGPK-----ESKETFDIITP----- 1227
Qy 822 GAFSDCLTIVBFLHSFGKVLGDFPAKDVPSLGLVQEGLLCQGSLSIGEVDLLVRLKAL 881
Db 1228 ---SD---VAEPLH---ETIVSEPAE---IQSEEEIEAQ---GEVDKLLFR----- 1264
Qy 882 HDGFPSPYCSQSLKITGEKVSIPILTRDNVNSIILCFLMAYGVXVAPALCDRLRTPFOAQP 941
Db 1265 -----SDTLQ--ITDLGVSGA-----REEFVETCP 1288
Qy 942 QOKAAVLAFPVHELNGSTLIINEIDKTLNESSSYKKNKWIIVEGRLRLKTVLAKTGRSE 1001
Db 1289 EHKGVIBSVVTIEDDTFVVQTTTD-----EG-----ESGHS 1321
Qy 1002 VEMGRPECLGRRRRSRIMEETSGMEEBEEESIAAVPGRRRRGRDGEVDATASSTPELER 1061

Db 1322 VRFAALEQPEVERRRSPHDEEFVEEABAO---ABP-----KDGSEAPAS--PEREE 1371
Qy 1062 QIEKLSKRQLF--FRKLLHSSQMLRAVSLGQDRYRRRYWVLPYLAFIVGTEGNLYPE 1119
Db 1372 VALSEYKTETVDYKDETTIDDSIMDADSL-----WV-----DTQDDRSIMTE 1415
Qy 1120 --EVIKKTDSLKVAHAASLNPALFSMKWELAGSNVTTASSPARARSRLKTKPGMQPRH 1177
Db 1416 QLETIPKEKAKEARRSLE-----KHKKEKPKTGRGRISTPE 1455
Qy 1178 FK-----SPVGGDSBPQOAL--OPEAOLHVPAPQ---OPOLQLOLQSHKGF 1221
Db 1456 RKVAKKSTVSRRDEVRKKAIVYKKAELAKTEVOAHSPSRFILKPAIKYTRPHLSCV 1515
Qy 1222 EQEGSPLSGOQHDLSQAFSLWLSQTSQSHSSLLSSSVLTDPSSPGKLDPAPOPPREP 1281
Db 1516 KR-----KITGDCSILQSA-----GGESALAPSVFKQAKDK 1546
Qy 1282 EPDEAESPDLOAFWNISAQWPCNAAPTPPLAVSEDOP-----TPSPQOLASSK 1331
Db 1547 VSDGVTKSPEKRS-----SLPRPSSILP-PRRGVSGDRDENSFSLNSSISSARRTRSE 1600
Qy 1332 PMNRSAAANPCSPVQFSSTPLA-GLAP---KERAGDPG--EMQSPGTGLGOKRGRPPS 1385
Db 1601 FIRRAGKSGTSTPTTGGTATPGTPPSSYSRKTPGTGTPSPRTPHTPGTKGAILVPS 1660
Qy 1386 KFFKQMEQRYLTQTAQVPPMCMGMMWI--PDPEMLDAMLK-----ALHPRGIREK 1436
Db 1661 -----EKKVAIIRTPKSPATPKQLRLINQPLDKNVSKIGSTDNIKYQPKGGVR 1713
Qy 1437 ALHK-----HLNKHDFLOEVCLRPSADP 1460
Db 1714 ILNKKIDFSKVQSRGSKDNIKHSAGGNVQIVTKKIDLSHVTSCGSLKNIRHRPGGR 1773
Qy 1461 IFEPQLPAFOEGINSWSPKETYETDLAVLQWVELEQRYVMSDLQIRGTCSPDSTR 1520
Db 1774 V-----KIESVKLDFEKAQAKVGSLD---NAHHVPGGGNVK 1807
Qy 1521 ED---LAYCEHLS---DSQEDITWRPGREGIAPORTKTNPLD-----LAVMRLLALEQ 1568
Db 1808 IDSQKLNFRHAKARVDHGABIIITQSPGRSSVASPRRLSNVSSGSLNLLSPQLATLAE 1867
Qy 1569 NVKRRYLREPL 1579
Db 1868 DVTRALAKQGL 1878

RESULT 46
US-09-949-016-8903
; Sequence 8903, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIORITY FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8903
; LENGTH: 1878
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-8903

Query Match 2.4%; Score 239.5; DB 4; Length 1878;

Best Local Similarity 19.5%; Pred. No. 5.3e-09;
Matches 318; Conservative 208; Mismatches 620; Indels 485; Gaps 74;
Qy 172 DSFPDQNFEEVCSGIHPDEAAEKEMT-----SVAE-----NGTGLVCSL 211
Db 510 DSFPVSLVLEQAVTDSAMTSKLEKANTFSALEIKSSIQELFEMRVDDKDKIEGVGAATSA 569
Qy 212 ELEXPQELKCMGNGSVSPVESLHQEVSVLPVPTVSCLDPPSHLPDQLEDTPILSBD 271
Db 570 ELDMPFYEDK-----SGMSKYFETSALEKATKSI---EPGSDYYELSDT---RESV 615
Qy 272 LEPNLSLAPVPSGGLYIGDIDTELGAEDKLPYDSDSPVIALDCPSLNNAATFSLADDS 331
Db 616 HESIDTWSPMHKG-----DKEPQTGKESQP---SPPAQBAQYSTLQAQSPSDLPPEPS 666
Qy 332 QTSSTISASPTSPVVLGE-----SVLQDNSFDLNGSDAEQESMET----- 372
Db 667 SPQBRMTT--IDPKVYGEKRDHLHKNKDDLTLSRLSLGLGGRSAIEQSRMSINLPMSCIDS 724
Qy 373 -----QSSDPPP-----SLTOPAPQSSSTIQLHPATSPAVS-----PTTSPAVSLVVS 415
Db 725 IALGFNFGRGHDLSPASDILNTSGMDEGDDVLPATTPALEKAPCFPVESKEEQIEK 784
Qy 416 PAASPEISPEV---CPAA-----STVSPAVFSVSPASS-AVLPAVSLVPLTAS 462
Db 785 VKATGEESTQABIESCFPLAKOFYKNGTVNAPDPEMLDLAGTRSLASVSADAEVARR 844
Qy 463 VTSPKASPVTSPPAAAPTASPAKDV---SSFLE-----TTADVEEITG 503
Db 845 KSVSETVEDSRTRLGDPVTDENHVIKTDLSQLEDLGYCVFNKVTVPPLSPVQSENLSG 904
Qy 504 E-GLTASGSGVMRRRIATP---EEVRLPLQHGWRREVRIKKGSHRMQGETWTYGPCGR 559
Db 905 ESGTFYEGTDDKVRDLATDLLEVLAAAGRVKDFSVDK-----EASAHISGDKSLG 959
Qy 560 MKQPPPEVIK-----YLSRNLVHSVRREHSPSPRMPVGDFFERDTPGELQWVQLSAB 612
Db 960 SKFPDQEKANDRLDVTLEKSEEHADSKHAKEETEE-AGDEIETFGI--GVTYEQALAK 1015
Qy 613 B--IPSRQAITKGRPRNTEKAKTEKVPKVRG-----RGRPPKVKITEL-----LN 659
Db 1016 DLSIPTDASSKAEGLSSVPEIAEVEPSKKVQGLDFAVQGG--LDVKISDFGQWASGLN 1074
Qy 660 KTDNRPLK- KLEAQETLNEEDKAK-----IAKSKKQMRQKVQRGECLETTIOGARNK 710
Db 1075 IDRRATELKLKLEATQDWTTPSSKAPQADAFMGVESHMKEGTVKSETEV-----K 1124
Qy 711 RKQETKSLKHKAAXKAEKEGKTK---OEKUK-EKVKEKKEKVKMKKEEVTAKK- 765
Db 1125 EKVAKPDVLVHQEAVDKEESYESSGHESLTMESLKADEGKETSPESLSIQDEIAVKLSV 1184
Qy 766 ---PACKADKTLATQRRLEERQKQMLLEBMKPTEDMCLTDHQPDPSPRGLTLP 821
Db 1185 EIPCPAVSEADLAT-----DERADVQ--EPIQGPKE-----ESKETPIDISITP----- 1227
Qy 822 GAFSDCLTIVEFLHSFGKGLGFDPAKDVPSLGVQLQEGLLCGSDSLGEVQDILLRLLKAAL 881
Db 1228 ---SD---VAEPLH---ETIVSEPAE---IQSEEEIEAQ---GEYDKLLFR----- 1264
Qy 882 HDGFPSPCYOSKLTGLKGVSEIPLTRDNVNSILRCFLMAYGVXPALCDRLRLTQFPQAOPP 941
Db 1265 -----SDTLQ--ITDLGVSGA-----REEFVETCP 1288
Qy 942 QOKAAVLAFPVHGLNGSTLIINEIDKTLSSNYSRKKNKWIVEGRLRLTLVAKRTGRSE 1001
Db 1289 EHKGVIESVVTIEDDFITVVQTTD-----EG-----ESGHS 1321
Qy 1002 VEMGRPECLGRRRSSRIMEETSCWEEBEEESIAAVPGRGRDRGEVDATASSTPELER 1061
Db 1322 VRFAALEQPEVERRRSPHDEEFVEEABAO---ABP-----KDGSEAPAS--PEREE 1371
Qy 1062 QIEKLSKRQLF--FRKLLHSSQMLRAVSLGQDRYRRRYWVLPYLAFIVGTEGNLYPE 1119

Qy	796	----	PTDMCLTDHQ-----	PLPDSRVPGLT--LPSGAFSDCLTIVIFEL-----	HS	836
		:	:	:	:	:
Db	1111	TPGSEE-SMTDEKTTKIACKSPPESDVPTSTKQWPKRSLRKADVEEFALRKLTPS	1169			
Qy	837	FGKVL-----	GDPDAKDVP5-LGVLEGILLCOGDSLGEVQDILLVLLKA-ALHD-PGF-	886		
Db	1170	AGKAMLTPKAGGD-EKDIKAFMTGPVKQDLAGTLPGRKQLQTPKEKAQALDLGFK	1228			
Qy	887	-----	PSYCOSLKILGEKVBSEIPL-----	TRDNVSEILRCFL	918	
Db	1229	ELFQTPGHTHEELVAAG-KTKIKPCD5SQSDPVDPTSTKQPKRSIRKADVEGELLACRN	1287			
Qy	919	MAYGVXA-----				
Db	1288	LMP5AGKAMHTPKPSVGEEKDIIIFVGTGPVKQLDLTENLTGSKRRPQTPKEEAQAL-	1343			
Qy	953	HELNGSTLIINEIDKLTLESMSYRKNKIWVE-----	GRLRRLKTVLAKRTGRSE	1001		
Db	1344	EDLTGFKELQTPGHTTEEAAGAATTQWPCESPPE5ADTPTSTRQPKTPLKRDVQKE	1403			
Qy	1002	V-----		EMGRPECEIGRRSSSRIMETS	1025	
Db	1404	LSALKKLTQTSGETHTTDKVPGEDK5INAFRETAQKQLPAA5VTGSKRHPKTKEKAQP	1463			
Qy	1026	MEE-----	BEESIIA5VPCRRRRRGEVDATASSIPELERQIEKL	1066		
Db	1464	LEDLAGWKELFQTPVCTDKPTTHEKTTKIAC-----	RSQPD-PVDPTPTSSKPSQKRSRLKV	1518		
Qy	1067	SKROLFP--RK-----	KLHSSQMLRAVSLGODRY-----		1094	
Db	1519	DVEEFPALRKRTPSACKAMHTPK--PAVGEKNIYAFMGTGPVKQLDLTENLTGSKRRLQ	1576			
Qy	1095	--RRRYWVLYLAGIFVE--GTGNLVPEEVIKKTDSLKVAHA5-----	LNPALFSMK	1145		
Db	1577	TPKEKAQALDLGAG-FKELFQTRGH--TEESMTND-KTAKVACKSSQPDLDKNP5ASSKRR	1632			
Qy	1146	ME-----	LAGSNTTASSPARAR5RPLKTGPFQWPHFK5PVRGQD5EQ-----	1189		
Db	1633	LKTSGLGKVGEELLAVGKLTQTSGETHTHTBPTGDKSMKAFM5PKQILDS5AASLTG	1692			
Qy	1190	POAQLQ-PEAQLHVPAPQPOLQI-QLOSH--KGFL5EQGSPL5L5GOSQHDL-----	1237			
Db	1693	SKQLRTPCK5EVPEDLAGFIELFQTPSHTK5SMTNEKTTKVSYRASQDPLVDTTSSK	1752			
Qy	1238	-----	SOSAF5LWSLQTSQSH5SL55VLT5D5SPG-----	KLDP	1272	
Db	1753	POPKR5LRKADTBEEFLAFKQTP5-----	AGKAMHTPKPAVGE5EQINTPLGTPVKQLDQ	1808		
Qy	1273	AP5OPPE---EPEPDEA55PDLOAF5WFI5AQMPCNAAPT5PLAV5EQDPT5PQOLA	1328			
Db	1809	PGMLPG5NRRLQTRKEKAQALEELTG--FREL5QTPCTDNP5TDEKTTKILCK5SQSDP	1866			
Qy	1329	5SKPMN---RPSA5NCP5VQ-----	F55T5PLAGLAPKRRAGDPG5M5Q5PT5GLQ5P--	1377		
Db	1867	ADTPTNTKQPKR5LKKADVEE5EFLAFKLT5PSAGKAMHTPKPAVGE5EQINT5FVGT5PVE	1926			
Qy	1378	-----	KRRGR5P55K5P5Q5MEQ-----	RYLTQ-----	LTAQ5PVP5CM5CG5WW	1414
Db	1927	KL5DL5LGNL5G5KRR5PQTP5PEKAKALEDL5AGFKELFQTPGHT5E55MTDDK5T5E55CK5PQ-	1985			
Qy	1415	IPDPEMLDAMLK-----	ALHPRG5IREKAL-----			
Db	1986	-PDPVKTPT55KQBLK5L5GKV5GVE5VL5PVGKLTQTS5GKTTQHT5ETAGD5GK5I5KAFKE	2044			
Qy	1457	SADP5I5EP5QL5PAF5Q5E5G5M55W--5PKE5TY5ET5DL5VL5QW5EE5E5QR5VIM5DQ5LQ5IRG5W5TCP	1514			
Db	2045	5AKQMLDPAN--YGT5G5MER5W5RTP5KEEA--Q55LEDLAGFKELFQ-----				2084
Qy	1515	5P5D5TRED5LAY5CEHL5D55Q5D5I5TW5RG5R5GL-----	APQ5RTN5PL---	DL5W5R5LA5LE	1567	
Db	2085	TPDHT5E55T-----	DDKTT5KIACK5PP5555MDTPT5TR5R5PKT5PL5KRD5I5VEEL5SALK	2138		

Qy	1568	Q	-----NVKRYRLREPLWPTTHEVLEKALLSTPNG-----	1597
Db	2139	Q	LTQTTHTDKVPGDEKGINVFRETAKQKDLPAASVTSKGRQRPTRPKGAQPLEDLAGLK	2198
Qy	1598	---	APEGTTTEISYEITPRIRIWRQTLQRCS	1626
Db	2199	ELFQTPICTDKPTTHEKTKI	-----ACRS	2223
RESULT 48				
US-09-976-594-22				
; Sequence 22, Application US/09976594				
; Patent No. 6673549				
; GENERAL INFORMATION:				
; APPLICANT: Furness, Michael				
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS				
; FILE REFERENCE: PA-0041 US				
; CURRENT APPLICATION NUMBER: US/09/976,594				
; CURRENT FILING DATE: 2001-10-12				
; PRIOR APPLICATION NUMBER: 60/240,409				
; PRIOR FILING DATE: 2000-10-12				
; NUMBER OF SEQ ID NOS: 1143				
; SOFTWARE: PERL Program				
; SEQ ID NO 22				
; LENGTH: 3256				
; TYPE: PRT				
; ORGANISM: Homo sapiens				
; FEATURE:				
; NAME/KEY: misc feature				
; OTHER INFORMATION: incyte ID No. 6673549 27001132CD1				
US-09-976-594-22				
Query Match 2.4%; Score 236.5; DB 4; Length 3256;				
Best Local Similarity 19.1%; Pred. No. 2.3e-08;				
Matches 442; Conservative 256; Mismatches 788; Indels 827; Gaps 114;				
Qy	1	MEMEANFANDHFNFTGLPPAPASGLKPSP	-----SSGGLYNGSPM	43
Db	51	IEIHEQEAAILH-NFSSNTPTQVNGSVIDEPVRLKHGDVITIIDRSFRYENESLOSGRKST	109	
Qy	44	NFPOQ-----GKSLGNDVNVNGLSTVSHT-TTSGILNSAPH-----SSSTSH	84	
Db	110	EFRKIREQPARVRVRSFSSDPDEKAQDSKAVSKITEGKVSQNPQVHIKNVKEDSTAD	169	
Qy	85	LHPSVAYDCIMNYQVPSANPNSLNDP-----PLJSQFSG-GQYPLNGILGG	132	
Db	170	DSKDSVAQGTNTNVHSSHAGRNGRNAADPIGDPKEISSVKLVSRYGELKSVPTTQCILDN	229	
Qy	133	SRQSSSHNTNLNLAGSKQFWANGTHSPMGLNFDSEQELYDSFPDQNFEEVC--SGIHPDE	190	
Db	230	SKNESP-----FWKLYESVKKELDVKSQK-----ENVLQYCRKSGLOTDY	270	
Qy	191	AAEKEMTSVVAENGTGLVCSLE-----LBEXQPELKMCGYNGSVSPVESLHQE	238	
Db	271	ATEKESADGL-QGETQLLVSRKSRPKSGSGSHA VAEPA SPEQELDKNGKGRDVE----	324	
Qy	239	VSVLVDPPTVSCLDPPHLPDQLEDTPI-LSEDSLEP-----FNSLADEPV----	SG	285
Db	325	-SVQTPSKAVGA-SFPLYEPKMK-TPVQSYQQQNSPQKHKNKDLTYTTGRRRESYNLGKSE	381	
Qy	286	GLVGIDDT---ELMGABDKLPL-XDSPVIALDCPSLNNATAFSLAADDS--QTSTSIFA	339	
Db	382	GFKAGDKTLTPRKLTSTRNRTPAKVEDAADSATKPELSSKTRGSIPTDVEVLPTETIHN	441	
Qy	340	SP-----TSPPVLGESVLQD-----NSFDLNNNGSDAEQE-----	368	
Db	442	EPFLTLMLTQVERKI QKDSLSKPEKLTGTAGCMCSGLPGLSSVDINNFGDSINESGIP	501	
Qy	369	-----EMETQSDPFPPLI---TQAPADOSTIQLHPATSPAVSPPTTSFAVLVV	414	
Db	502	KRRRVSGGHLRPELFDENLPPTNPLKRGGAFTKRLVMH-----TPVLKKII	551	

Qy 415 SPAASP-----EISPEVCAASTVSVSPAVFVSPASSAVLPAVSVLEVLFTASVTSP 466
Db 552 KEQPFQSGKQSGSEIHVEV-KAQLVISP-----PAPS-----PRKTPVASD 593
Qy 467 -----KASPVTSAPAAPPTASPANKOVSSPLETTADVEEITGGLTASGSDVWRRRIA 520
Db 594 QRRRSCKTAPASSKSSOTEVPRKGERVATCLQKRVSISSQHDILQW-----ICSKRRS 648
Qy 521 TPEEVRLPLOGHMRREVRKKGSHRWQETWYGP---CGKRM-----KQPEVIKYL 570
Db 649 GASEANLIVAKSMADVVKL--GAKQOTKVIKHGQFSMNKRQRRPATPKPKPGEVHSQF 706
Qy 571 SRNLVHS-----VRREHFSFPRMPVGDF--EERDTPBGLQWV---QLSAEEI 614
Db 707 STGHANSPTIIIGKAHTEKVHVPARPYRVNLNFINQKXDFKEDLSGIAEMFTPVKEQ 766
Qy 615 P-----SRIQAITGKRGR-----627
Db 767 POLTSTCHIAISNENLLGKQFGTDSGEEPLPTSBSFGGNVFFSAQNAAKQSPDKCSA 826
Qy 628 -----PRTEKAKTYEVKVKVGRGRPPKVKITELLNK-----TDRNP 665
Db 827 SPPLRQCIRENGNAVKTPTNRYKMTSLET-KTSDTETEPSKTVST--VNRSGRSTFRN 883
Qy 666 LKLEAOTLNEEDKAKIAKSKMKRQKVRQEGCLTTIQOARNRKO-----ETK 716
Db 884 IQKLPV-ESKSEENTWELVEILK-----RGQATILQORREGEMKEIERPPFTYKNI 936
Qy 717 SLKHKEAKKS-----XAE 730
Db 937 ELKENDERKAMKRSRTWGQCAPMSDLTLKSLPDTELMKDRTARGQNLLOTQDHAKAPK 996
Qy 731 KEKSK-TK--QEKLEKVKREKEKVKMKE-----KEV-----TKA 764
Db 997 SEKGKITMPQCSLQPEPINTPTHTKQOLKASLGKVGKVEALLAVGKFTRTSGTHTHR 1056
Qy 765 KPACKADKTLATQRLEERQOQMIIE-----MKK-----795
Db 1057 EPAGDG-KSI---RTFKESPKQ--ILDPAARVTGKMKWPRTPKEASQSLDLAGFKELFQ 1110
Qy 796 ---PTDMCLFDHQ-----PLPDFSRVPGLT--LPSGAFSDCLTIVEFL-----HS 836
Db 1111 TPGPSEE-SMTDEKTTKIACKSPPPESVDPTSTKQWPKRSRKADVEEFLALRKLTPS 1169
Qy 837 FGKVL-----GFDPKADVPS-LGVLOEGLLCOGDSIGEVDLLVRLKA-ALHD-PGF- 886
Db 1170 AGKAWLTPKAPAGD-EKDIFAMGTPVQKLDLAGTLFGSKRQLOTPEKKAQLEDLAGFK 1228
Qy 887 -----PSYCQSLKITLGEKVSBIPL-----TRDNVSEILRCFL 918
Db 1229 ELFTQPGHTELVAG-KTKIIPCDSPQSDPVDPTPTSTKQPKRSIRKADVEGELLACRN 1287
Qy 919 MAYGVXPA-----LCDRLRTQFPQAPQPOQKAAVLAPV 952
Db 1288 LMPISAGKAMHTPKPSVGEEKDIIIFVGPVQKLDLTENLTGSKRRPQTPKEEAQAL---- 1343
Qy 953 HELNGSTLIIINEIDKTLIESMSYRNKWIPE-----GRLRLKTVIAKRTGSE 1001
Db 1344 EDLTGFKELFQTPGHTTEAAGKTTKMPCESSPPESADTPTSTRQPKPTPLEKRDVQKE 1403
Qy 1002 V-----EMGRPEECIGRRSSRSMETSG 1025
Db 1404 LSAUKLUTQTSGETHTDQVPGGEDKSNAPRETAKQKLDPAASVTGSKRHPTKEKAQP 1463
Qy 1026 MEE-----EEEEESIAAVPGRGRGRDGEVDATASIPPELEROIEKL 1066
Db 1464 LEDLAGWKELFQTPVCTDKPTTHEKTTKIAC---RSQPD-PVDTPTSKPKQSKSLRKV 1518
Qy 1067 SKRLQFP--RK-----KLLHSSQMLRAVSLGQDRY-----1094
Db 1519 DVEEEFALRKRTPSAGKAMHTPK--PAVSGEKNYIAYFMGTPVQKLDLTENLTGSKRRLQ 1576
Qy 1095 --RRRYWVLPYLAGIFVE--GTEGNLVPEEVIKKETDSLKVAHAAS-----LNPALFSMK 1145

Db 1577 TPKEKAQLEDLAG-FKELFQTRGH--TEESMTND-KTAKVACKSSQDLDKNPASSKRR 1632
Qy 1146 ME-----LAGSNTTASSPARARSLPKTKGFGWQPRHFKSPVRQDSEQ---- 1189
Db 1633 LKTSLGKVGKVEILLAVGKLTQTSGETTHTEPTGDKGKMAFMESPQKILDSAASLTG 1692
Qy 1190 PQAQLO-PEAQLHVPAQPOQLQ-QLQSH--KGFLQEGSPLSLGOSQHDL----- 1237
Db 1693 SKQLRTPKGRSEVPEDLAGFIELFQTPSHTKESMTNEKTKVSYRASQDLDVDTPTSSK 1752
Qy 1238 -----SOSAFLSWLSQTSQSHSSLLSSSVLTPDSSPG-----KLDLP 1272
Db 1753 PQPKRSRKADTEEEFLAFRKQTPS-----AGKAMHTPKPAVGEEKDINTFLGTVPQKLDQ 1808
Qy 1273 APSOPPE-----EPDPDAESPDLOAQFNFNISAQPCNAAPTPPLAVSEDOPTSPQOLA 1328
Db 1809 PGNLPGSNRRLQTRKEKAQALELTG--FRELFTQPTCTDNPTTDEKTTKKILCKSPQSDP 1866
Qy 1329 SSKPMN---BPSAANPCSPVQ-----FSSTPLAGLAPKRRAGDPGEMPOSPTGLGQP-- 1377
Db 1867 ADTPTNTKQRPKRSLLKADVEEFLAPRKLTSPSAGKAMHTPKAAVGEKDKINTFVGTPVE 1926
Qy 1378 -----KRRGRPPSKFFKQMEQ-----RYLTQ-----LTAQPVPPMCSGWW 1414
Db 1927 KLDDLGNLPGSKRRPQTPKEKAKALEDLAGFKELFQTPGHTESMTDDKITEVCSKSPQ- 1985
Qy 1415 IPDPEMLDAMLK-----ALHPRGIREKAL-----HKHLNKHQDFLQF-----VCLRP 1456
Db 1986 -PDPVKTPTSSKQRLKISLGKVGKVEVLVPGKLTQTSKGTQTHRETAGDGKSIKAFKE 2044
Qy 1457 SADPIFEPRQLPAPQEGISMW--SPKEKTYETDLAVLQWVEELQORVIMSDLQIRGTCP 1514
Db 2045 SAKOMLDPAN---YGTGNERWPKTPKEEA--QSLEDLAGFKELFQ-----2084
Qy 1515 SPQSTREDLAYCEHLSDSQEDITWRGFGREGL-----APQKRTTNPL---DLAVMRLAAL 1567
Db 2085 TPDHTEESTT-----DDKTTKIACKSPPPESMDPTTSTRRRPKTPLGKRDI-VEELSALK 2138
Qy 1568 Q-----NVKRYLREPLWPHWVLEKALLSTPNQ-----1597
Db 2139 QLTQTTHTDKVPGEDKGINVFRETAKQKLDPAASVTGSKRQRTPKKAQPLEDLAGLK 2198
Qy 1598 ----APSGTTTEISYEITPRIRIWRQTLQRCRS 1626
Db 2199 ELFTQPTICTDKPTTHEKTTKI-----ACRS 2223
RESULT 49
US-09-919-039-21
; Sequence 21, Application US/09919039
; Patent No. 6727066
; GENERAL INFORMATION:
; APPLICANT: Kaser, Matthew R.
; TITLE OF INVENTION: GENES EXPRESSED IN TREATED HUMAN C3A LIVER CELL CULTURES
; FILE REFERENCE: PA-0035 US
; CURRENT APPLICATION NUMBER: US/09/919,039
; CURRENT FILING DATE: 2002-09-09
; PRIOR APPLICATION NUMBER: 60/222,113
; PRIOR FILING DATE: 2000-07-28
; NUMBER OF SEQ ID NOS: 401
; SOFTWARE: PERL Program
; SEQ ID NO 21
; LENGTH: 3256
; TYPE: PRN
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6727066 2700132CD1
US-09-919-039-21
Query Match 2.4%; Score 236.5; DB 4; Length 3256;
Best Local Similarity 19.1%; Pred. No. 2.3e-08;

Matches 442; Conservative 256; Mismatches 788; Indels 827; Gaps 114;

```
QY 1 MEMEANEANDHFNFTGLPPAPAAAGLXPS-----SSGELYTNGSPM 43
D5 111HEQEAILH-NFSGTNPQVNGSVIDEPVRLKHGDVITIIDRSFRYENESLQGRKST 109
QY 44 NPPQO-----GKSLGDNVNGLSVTSHT-TTSGILNSAPH-----SSSTSH 84
D5 110EFPRKIREQEPARRVSRFSFSDPDDEKAQDSKAYSKITEGKVGSNPQVHIKNVKEDSTAD 169
QY 85 LHPHVSAYDCLWNYSOYPSANPGSNLKP-----PLLSQFSG-GQYPLNGILGG 132
D5 170 DSKSVAOGTTNVHSEHAGNRNAADPIISGDFKEISSVKLVSRGELKSVPTTQCLDN 229
QY 133 SRQSPSHNTNLRAGSKFWANGTHSPMGLNFDSELYDSPDQNFEEVC--SGIHPDE 190
D5 230 SKKNESP-----FWKLYESVKELDVKSQ-----ENVLYCRKSLQDNY 270
QY 191 AAEKEMTSVAENGTLVCSLE-----LEEXQPELKMCGYNGSVPSVESLHQE 238
D5 271 ATEKESADGL-QGETQLLVSRKSRPKSGSGHAVAEPASPEQELDONKKGKRDVE----- 324
QY 239 VSVLPDPTVSCLDPSHLPOLEDPTI-LSEDSLEP-----FNSLAPPEPV-----SG 285
D5 335 -SVQTPSKAVGA-SFPLYEPKMK-TPVOYSOQQNSPQKHKNKDYTTGRRESVNLKSE 381
QY 286 GLYGIDDT-----ELMGAEDKPL-XDSPVISALDCPSLNATAFSLADD--QTSTSI 339
D5 382 GFKAGDKILTFRKLSRTRTPAKVEDAADSATKPNLSKRTGSIPTDVEVLPTETELHN 441
QY 340 SP-----TSPVLGESVLD-----NSFDLNGSDAEQE----- 368
D5 442 EPFLTLWLTVVERKIQDLSLXPEKLGTTAGCMCSGLPGLSVDINNFGDSINESEGIPL 501
QY 369 -----EMETQSSDFPPL--TQPADQSSITQLHPATSPAVSPPTSPAVSLV 414
D5 502 KRRVSGFGLHURPELFDNLPTPLKRGEPATKRKSLVMH-----TPPVKKII 551
QY 415 SPAASP-----BISPEVCPAASTVSPVAFVSVPASSAVLPVAVSLEVPLTASVTSP 466
D5 552 KEOPQSKQSGSEIHVEV-KAQLSVISP-----PAPS-----PRKTPVASD 593
QY 467 -----KASPVTSPPAAFTTASPAKDVSSSLETTADVEEITGELTASGSDVNMRR 520
D5 594 QRRRSCKTAPASSKSQTEVPKRGERVATCLOKRVISRSQHDILQ-----ICSKRS 648
QY 521 TPEEVRPLQHCWRREVRKKGSHRWQGETWYGP--CGKEM-----KQPEVIKYL 570
D5 649 GASEANLIVAKSWADVVKL--GAKQTQTKVHKHQPQRSMNKRQRRPATPKXPVGVHSQF 706
QY 571 SRNLVHS-----VRRHFSFSPRMPVGDFE--EERDTPEGLQWV-----OLSABE 614
D5 707 STGHANSCTIIIGKAHTEKHVHPARVRLNLFISNQKDFKEDLSGIAEMFTPVKEQ 766
QY 615 P-----SRIQAITGKRGR----- 627
D5 767 PQLTSTCHIAISNENLLGKQFGTDSGEPLLTSESFGNVFPFAQAQPSDKCSA 826
QY 628 -----PRNTEKAKTEVPKVKRGGRPPKVKITELLNK-----TDNRP 665
D5 827 SPPLRRQCIRENGNVAKTPRNTYKMTSLET-KTSDTETETPSKTSTVST--VNRSGRSTERN 883
QY 666 LKLEAQETLNEEDKAKLAKSKKMKQKVORCECLTTIQOARNRKO-----ETK 716
D5 884 YQKLPV-EKSEETWEIVECILK-----RGQKATLIQORREGKMEIERPFYKENI 936
QY 717 SLKHKEAKKKS-----XAE 730
D5 937 ELKENDERKWKAMKRSRTGQKCAPMSDLTDLKSLPDTMLMKDTARGONLLQTDHAKAP 996
QY 731 KEKGK-TK--QEKUKEKVKREKKEKVKKE-----KEEV-----TKA 764
D5 997 SEKGKITKMPQCSQLOPEINTPTHTKQOLKASLGKVGKVELLAVGKFTRTSGETHTHR 1056
```

```
QY 765 KPACKADKTLATQRRLEERQKQMILEE-----MKX----- 795
D5 1057 EPAGDG-KSI---RTFKESPKQ--ILDPAARVTGMKWPRTPKBEAQSLDLAGFKELFQ 1110
QY 796 ---PTEDCLTDHQ-----PLPDFSRVPCLT--LPSGAFSDCLTIVEEL-----HS 836
D5 1111 TPGSEB-SMTDEKTTIACKSPFESVDTTSTKQMPKGRSLRKADVEEELALUKLTPS 1169
QY 837 FGKVL-----GFDPKADVPS-LGVLOEGLLCQSDSLGEVDLLVRLKA-ALHD-PGF- 886
D5 1170 AGKAMLTPKAGGD-EKDIKAFMTGPVQKLDLAGTLPCKEQLQTPKEKAQLEDLAGFK 1228
QY 887 ---PSYCOSLKILGEKVSIEPL-----TRDNVSEILRCLF 918
D5 1229 ELFTQPGHTEBELVAAG-KTKIPCDSPQSDPDVDTSTKQPKRSIRKADVEGELLACRN 1287
QY 919 MAYGVXPA-----LCDRLRTPQFQAOPOQKAAVLAFPV 952
D5 1288 LMPGAGKAMHTPKPSVGEEKDIIIFVGTVPQKLDLTENLTGSKRRPQTPKEAQAL----- 1343
QY 953 HELNGSTLIINEIDKLTLESMSYRNKNWIVE-----GLRLRLKTVLAKRTGRSE 1001
D5 1344 EDLTGFKELFOTPGHTEAVAGKTTKMPCESSPPESADTPTSTRRQKTPLEKRDVQKE 1403
QY 1002 V-----EMGRPEBCLGRRRSRIMEETSG 1025
D5 1404 LSALKKLTQTSGETTHTDKVPGEDKSIINAFRETAOKOKLDPAA SVTSGKRHPKTKKAQP 1463
QY 1026 MEE-----EEEEESIAAVPGRGRRDGEVDATASSIPELEROIEKL 1066
D5 1464 LEDLAGMKELQTPVCTDKPTTHBKTAKIAC-----RSQPD-PVDTPPTSSKPSKRSRKV 1518
QY 1067 SKROLFF-RX-----KLLHSSQMLRAVSLGQDQY----- 1094
D5 1519 DVEEFALRKRTPSAGKAMHTPK--PAVSEKNIYAFMGTPVQKLDLTENLTGSKRRLO 1576
QY 1095 --RRRYWVLYLAGIFVE--GTEGNLVPVEEVIKKETDSLKVAHAAS-----LNPAIFSMK 1145
D5 1577 TPKEKAQALEDLAG-FKELFQTRGH--TEESMTND-KTAKVACKSSQPDLDKNPASSKR 1632
QY 1146 ME-----LAGSNTTASSPARARSPLTKPGFMPQPRHFKSPVRQDSEQ----- 1189
D5 1633 LKTSLGKVGKVELLAVGKLTQTSGETHTHTPTGDKSMKAFMESQPLDLSAASLTG 1692
QY 1190 POAOLQ-PEAQLHVPAPQPOLQ-LQOSH--KGFLEQEGSPSLGSOHDL----- 1237
D5 1693 SKRQLRTPKSKSEVPEDLAGFIELFQTPSHTKESMTNEKTTKVSYRASQPDLDVPTSSK 1752
QY 1238 -----SQSAFLSWLSQTSKSHSLSSSVLTLPDSSPG-----KLDLP 1272
D5 1753 PQPKRSLRKADTEBEFLAFRKQTPS---AGKAMHTPKPAVGEEKDINTFLGTPVQKLDQ 1808
QY 1273 APSOPPE-----EPDPDEAESPDLOAFWFNISAOQPCNAATPPPLAVSEDOPTSPQOLA 1328
D5 1809 PGNLPGSNRRLQTRKEKAQALELTG--FRELFQTPCTDNPDTTDEKTTKILCKSPQSDP 1866
QY 1329 SSKPMW--RPSANPCSPVQ-----FSSSTPLAGLAPKRRAGDPGEMPSPTGLQGP-- 1377
D5 1867 ADTPTNTKQRPKRSKADVEEBFLAFKLTSPSAGKAMHTPKAAVAGEBKDINTFVGTPE 1926
QY 1378 -----KRRGRPPSKFFKQMEQ-----RYLTQ-----LTAQVPPEMCSGMMW 1414
D5 1927 KLDLLGNLPGSKRRRPQTPKEKAKALEDLAGFKELFQTPGHTEESMTDDKITEVSKSPQ- 1985
QY 1415 IPDEMLDAMUK-----ALHPRGIREKAL-----HKHLNKHDFLOE-----VCLRP 1456
D5 1986 -PDVKTPTSSKQRLKISLGKVGKVELVPGKLTQTSQTKTQTHRETAGDKSIKAFKE 2044
QY 1457 SADPIFEPRQLPAQEGIMSW--SPKEKTYETDVLQWVELEQVIMSDLQIRGWTCP 1514
D5 2045 SAKQMLDPAN---YGTGNERMPRTPKEEA--QSLDLAGFKELFQ----- 2084
```

```
Qy 1515 SPDSTREDLAYCEHLSQEDITWRGPGREGI-----APQRTTNPL-----DLAVMLAALE 1567
Db 2085 TPDHTEEST-----DDTKTKIACKSPPPESMDIPTSTRRPKPLGKEDI-VELSALK 2138
Qy 1568 Q-----NVKRRYLRLPLMPTHEVLEKALLSTPNG----- 1597
Db 2139 QLTQTTHTDKVPGBDCKGINVFRETAKQKLDPAASVTGSKRPRTPKGAQPLEDLGLK 2198
Qy 1598 ---APECTTTEISYEITPRINWQTLCRCS 1626
Db 2199 ELFOPTICTDKPTTHEKTTKI-----ACRS 2223

RESULT 50
US-09-949-016-11650
; Sequence 11650, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11650
; LENGTH: 1043
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-11650

Query Match 2.4%; Score 235; DB 4; Length 1043;
Best Local Similarity 19.8%; Pred. No. 4.6e-09;
Matches 264; Conservative 190; Mismatches 472; Indels 408; Gaps 59;

Qy 24 SGLKPPSPSGGLYTNNGSPMNPPOQKSLNGDVNNGLSVTSHTTSGILNSAPHSSSTS 83
Db 5 SGLKPPRGKK-----HSSPM-----GRT-----STCSASSRAV-----AASSKEGS 42
Qy 84 HLHPSVAIDCLMNYSOYPSANPSNL-----KDPILLSQFSGQYPLNGILGSGRQPSPS 140
Db 43 PLH-----KQSSGSPSSPAAAAAPEKPGKAAEV-GDDELGLDFVVG----- 82
Qy 141 HNTNLRAGSQKFWANGTHSPMGLNFDLSQELYDSFPDQNFEEVCSGIHPDEAAEKEMTSVV 200
Db 83 -----ERVWNGVK-----PGVQYLGTFQAPQWAGW 112
Qy 201 AENGTGLVCSLEBEXQPELKMCGYNGSV-----PSVESLHQEVSVLVPDPTVSLDDPS 255
Db 113 LDDPVG-----KNDGAVGVRYFPCPALQGIITRPSKLTQRTAGSGSDA 158
Qy 256 HLPQLEDTPILSDESLPEFNSLAPEVSGGLYGIDDTFELMGAEDKLPXDSPVISALDC 315
Db 159 HSVES-----LTAQLNSLHSGTATPPLTSRV-----IPLRESVLNSVKT 198
Qy 316 -----PSLNNATAFSLADDSTQTSIFASTSPVP-----LGES-----VLQDINSFD 358
Db 199 QNKGSGNLSDSGSKVGRGEKURLGRVLVGGTKTGWRYVGETDFAKGEWCGBELDPLG 258
Qy 359 LNNGSDAQEEMETQSSDFPSPLOPADQSTTQLHPATSPAVSPETTSFAVSLVSPAA 418
Db 259 KNDGAVA-----GTRYFCQCPKFGLPFAPHKVIRIGFESTSPAKAKTKRMAMGVSAETH 313
Qy 419 SPEISPEVCPAASTVSPAVSVSPASS-AVLPAVSLEVLPTASVTPSPKASPTVSPAAA 477
Db 314 SPS-----SSSISVSVAASSVGGSPRSGLLTETTSRVARKISGTTALQEA 360
```

```
Qy 478 FPTASPANKOVSSFLETTADVEEITGEGLTASGSDVMRRRIATPE-----EVRLPLQHG 533
Db 361 LK-----EQQHIEQLAE-----RDLERAEEVAKATHICEV----- 392
Qy 534 RREVRIKKGSHRMGOETWYTGPCGKRMKQFPEVIKYLRSNLVHSVRREHFSFSPMPVGD 593
Db 393 EKEITALLKAQHE-----QYVAEAEKQLR-----ARLLVESVRKEKVDLSNQL----- 436
Qy 594 FFEEDTPEGLOWVOLSABEIPSIQAIT-GKGRPRNTEKAKTYEKVVKVGRGRPPKV 652
Db 437 --EERRKVEDLQF--RVEE-----ESITKGDLEQTQLEHARIGELEQ----- 475
Qy 653 KITELINKTNRPLKLEAOETLNEEDKAKIAKSKKMKRQKVQREGCLTTTIQGOARNRK 712
Db 476 ---SLL-----LEKAQAEERLLR-----LADNR-----LTTVAEKSRVLQL 508
Qy 713 QETKSLKHKEAKKXSA-----EKEKGTKQEKLEKKEKREKKE 751
Db 509 EEEELTLARGEIEELQOCLLHSGPPPPDPDAAEILRLRERLLSASKEHQRESGVLRDKYE 568
Qy 752 KVMKEKEEVTKAPACK--ADKTLATORLEERQKQOMILEEMKKTEDMCLTDHQ-PL 808
Db 569 KALKAYQAEVDKLRANNEKYAQEVAGLKDKVQQAATSENMGMLMDNWKSLDLSADSHQKSL 628
Qy 809 PDSRVPLGLTIPSGAFSDCLTIVEFLHSFGKVLGDFPAKDVPSLGVLOEGLCCGD-SLG 867
Db 629 EDLK-----ATLNSGPGAQ-----OKEIGELKAVNEGIMHEHOLELG 665
Qy 868 EVQDLLVRLKAAALHDPFPFSCOSKILGEKVEIPLTRDNDVSEIILRCFIMAYGVAPAL 927
Db 666 NLQ-----AKHDLTAMHVKEKEALREKLQ-----AQEELAGLQRHWAQLEV-QAS 712
Qy 928 CDRLRTOFPQAPQPOQKAAVLAPFVHEL-----NGSTLIINEIDKTLESMSYRKWKW 980
Db 713 QHRLQLQ--EAQDQRDAEL---RVHELEKLDVEYRGAQAI-----BFLKEQISLAEKKM 763
Qy 981 IVEGLRLAKTVLAKRTGRSEVEMGRPEECILGRRSRI-----MEETSGMBEE- 1029
Db 764 LDYERLQ-----AEAQKQEVESLREKLLVAENRLOAVEALCSSQHTHMIESNDISEET 818
Qy 1030 -EEESIAAVPGRRRGRDGEVDATASSIPELERQIEKLSKRQLFRKK---LLHSSQMLR 1085
Db 819 IRTKETVEGLQDKLNKEDKEVTALTSTQTEMLRAQVSALESCKSGEKKVDALLKEKRLE 878
Qy 1086 A-----VSLGQDRVRRYRWVLPYLAGIFVEGTEGNLVPEEVIKKETDLSLKVAHAHSLPAL 1141
Db 879 AELETVS-----RKTHDASGQLV-----LISQELLRKE-----RSLN--- 910
Qy 1142 FSKMELAGSNTTASSPARARSRLPKTKPGFMOPRHFKSPVR-----GQDSQPOQAOL 1194
Db 911 -ELVLLLELRNHSFGPERDLSREVHKAEWRIKQKUKDDIRGLREKLTGLDKK----- 964
Qy 1195 QPEAQLHVPAPQPOLQLOLQSHKHGFLEQEGSPLSLQSQSHDL--SQSAFLSWLSQTSQSH 1252
Db 965 -----SLSQDRRYSLIDPSSAP-ELLRLQHQLMSTEDALRDALDQAOV 1007
Qy 1253 SLLSSSVLPDSS 1266
Db 1008 EKLMENARSCPDKA 1021
```

Search completed: March 9, 2005, 14:57:32
Job time : 100.043 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 9, 2005, 14:24:09 ; Search time 126.703 Seconds
(without alignments)
4877.138 Million cell updates/sec

Title: US-10-702-148-13
Perfect score: 9826
Sequence: 1 MEMEANDHFNFTGLPPA.....MRRFFESRWBEFYQKQANL 1878

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1391452 seqs, 329044822 residues

Total number of hits satisfying chosen parameters: 1391452

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 100%

Listing first 65 summaries

- Database : Published Applications AA:*
- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
 - 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
 - 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
 - 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
 - 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
 - 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
 - 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
 - 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
 - 9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
 - 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
 - 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
 - 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
 - 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
 - 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
 - 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
 - 16: /cgn2_6/ptodata/1/pubpaa/US10D_PUBCOMB.pep.*
 - 17: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
 - 18: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pep.*
 - 19: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
 - 20: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	9816	99.9	1878	9	US-09-839-479-13
2	9816	99.9	1878	15	US-10-376-537-13
3	9816	99.9	1878	15	US-10-702-148-13
4	9712	98.8	1876	9	US-09-839-479-70
5	9712	98.8	1876	15	US-10-376-537-71
6	9712	98.8	1876	15	US-10-702-148-70
7	9590.5	97.6	1873	13	US-10-087-192-666
8	6777.5	65.0	1586	13	US-10-087-192-663
9	2482.5	25.3	1972	9	US-09-839-479-21
10	2482.5	25.3	1972	15	US-10-376-537-21
11	2482.5	25.3	1972	15	US-10-702-148-21
12	2436	24.8	1969	9	US-09-839-479-71
13	2436	24.8	1969	15	US-10-376-537-72

ALIGNMENTS

RESULT 1

US-09-839-479-13
; Sequence 13, Application US/09839479
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042002
; CURRENT APPLICATION NUMBER: US/09/839,479
; PRIOR FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24

14	2436	24.8	1969	15	US-10-702-148-71	Sequence 71, Appl
15	1030	10.5	196	14	US-10-062-831-94	Sequence 94, Appl
16	1030	10.5	196	14	US-10-062-599-94	Sequence 94, Appl
17	875.5	8.9	175	15	US-09-839-479-64	Sequence 64, Appl
18	875.5	8.9	175	15	US-10-376-537-65	Sequence 65, Appl
19	875.5	8.9	175	15	US-10-702-148-64	Sequence 64, Appl
20	671.5	6.8	1674	9	US-09-839-479-1	Sequence 1, Appl
21	671.5	6.8	1674	15	US-10-376-537-1	Sequence 1, Appl
22	671.5	6.8	1674	15	US-10-702-148-1	Sequence 1, Appl
23	669	6.8	1673	9	US-09-839-479-69	Sequence 69, Appl
24	669	6.8	1673	15	US-10-376-537-70	Sequence 70, Appl
25	669	6.8	1673	15	US-10-702-148-69	Sequence 69, Appl
26	575	5.9	141	9	US-09-925-237-555	Sequence 555, App
27	521.5	5.3	1525	9	US-09-839-479-68	Sequence 68, Appl
28	521.5	5.3	1525	15	US-10-376-537-69	Sequence 69, Appl
29	521.5	5.3	1525	15	US-10-702-148-68	Sequence 68, Appl
30	521.5	5.3	1527	9	US-09-839-479-27	Sequence 27, Appl
31	521.5	5.3	1527	15	US-10-376-537-27	Sequence 27, Appl
32	521.5	5.3	1527	15	US-10-702-148-27	Sequence 27, Appl
33	512.5	5.2	1531	9	US-09-839-479-29	Sequence 29, Appl
34	512.5	5.2	1531	15	US-10-376-537-29	Sequence 29, Appl
35	512.5	5.2	1531	15	US-10-702-148-29	Sequence 29, Appl
36	472	4.8	176	9	US-09-839-479-65	Sequence 65, Appl
37	472	4.8	176	15	US-10-376-537-66	Sequence 65, Appl
38	472	4.8	176	15	US-10-702-148-65	Sequence 65, Appl
39	406.5	4.1	2781	15	US-10-263-929-122	Sequence 122, App
40	351	3.6	160	9	US-09-764-864-1165	Sequence 1165, Ap
41	308	3.1	59	9	US-09-839-479-51	Sequence 51, Appl
42	308	3.1	59	15	US-10-376-537-52	Sequence 52, Appl
43	308	3.1	59	15	US-10-702-148-51	Sequence 51, Appl
44	305	3.1	2803	15	US-10-415-187-5	Sequence 5, Appl
45	295.5	3.0	1855	16	US-10-437-963-144559	Sequence 144559,
46	288	2.9	2468	16	US-10-755-889-615	Sequence 615, App
47	280.5	2.9	2364	15	US-10-205-331-66	Sequence 66, Appl
48	273	2.8	45	9	US-09-839-479-61	Sequence 61, Appl
49	273	2.8	45	15	US-10-376-537-62	Sequence 62, Appl
50	273	2.8	45	15	US-10-702-148-61	Sequence 61, Appl
51	269	2.7	319	14	US-10-106-698-5173	Sequence 5173, Ap
52	267	2.7	5560	15	US-10-263-929-142	Sequence 142, App
53	258.5	2.6	4952	15	US-10-051-874-56	Sequence 56, Appl
54	258.5	2.6	5008	15	US-10-051-874-166	Sequence 166, App
55	258.5	2.6	5159	15	US-10-085-198-112	Sequence 112, App
56	258.5	2.6	5262	15	US-10-051-874-165	Sequence 165, App
57	250.5	2.5	3507	15	US-10-369-493-5784	Sequence 5784, Ap
58	250.5	2.5	5262	15	US-10-051-874-167	Sequence 167, App
59	250	2.5	134	9	US-09-764-864-1578	Sequence 1578, Ap
60	246	2.5	2701	14	US-10-171-311-83	Sequence 83, Appl
61	245.5	2.5	4322	16	US-10-437-963-104793	Sequence 104793,
62	244.5	2.5	6642	15	US-10-369-493-5013	Sequence 5013, Ap
63	242.5	2.5	1268	15	US-10-276-774-2166	Sequence 2166, Ap
64	241.5	2.5	1234	14	US-10-203-311A-6	Sequence 6, Appl
65	238.5	2.4	1788	11	US-09-964-956-60	Sequence 60, Appl

```
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 1878
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(1878)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-08-839-479-13

Query Match          99.9%; Score 9816; DB 9; Length 1878;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1878; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MEMEANEANDHFNFTGLPPAPAAAGLKPSPSGEGLYTNGSPMNPPOQCKSLNGDVNNG 60
Db 1 MEMEANEANDHFNFTGLPPAPAAAGLKPSPSGEGLYTNGSPMNPPOQCKSLNGDVNNG 60

Qy 61 LSTVSHTTTSGILNSAPHSSSTSHLHHPVAYDCLWNYSQYPSANPGSNLKDPPLLSQFS 120
Db 61 LSTVSHTTTSGILNSAPHSSSTSHLHHPVAYDCLWNYSQYPSANPGSNLKDPPLLSQFS 120

Qy 121 GGOYPLNGILGSGRQSPSPSHNTNLRAGSQKFWANGTHSPMGLNFDISOELYDSDPDQNF 180
Db 121 GGOYPLNGILGSGRQSPSPSHNTNLRAGSQKFWANGTHSPMGLNFDISOELYDSDPDQNF 180

Qy 181 EVCSGIHDEAAEKEMTGVAAENGTLVCSLELEKQPELKMCGVNGSVPSVSLHQEVS 240
Db 181 EVCSGIHDEAAEKEMTGVAAENGTLVCSLELEKQPELKMCGVNGSVPSVSLHQEVS 240

Qy 241 VLVPDPTVSCLDPPSHLPDQLEDTPILSEDSLEPPNSLAPFVSGGLYIGIDDTELMAED 300
Db 241 VLVPDPTVSCLDPPSHLPDQLEDTPILSEDSLEPPNSLAPFVSGGLYIGIDDTELMAED 300

Qy 301 KLPLXDSPIVIGALDCPSLNNATAPSLADDSQTSTSIFASPTSPVLGESVLQNSFDLN 360
Db 301 KLPLXDSPIVIGALDCPSLNNATAPSLADDSQTSTSIFASPTSPVLGESVLQNSFDLN 360

Qy 361 NGSDAEQEMETQSDFFPSLTQAPDQSSITQLHPATSPAVSPPTSPAVSLVSPAAASP 420
Db 361 NGSDAEQEMETQSDFFPSLTQAPDQSSITQLHPATSPAVSPPTSPAVSLVSPAAASP 420

Qy 421 EISPEVCPAASTVTVSPAVFVSPASSAVLPAVSLVPLTASVTSPKASPTVSPAAAPPT 480
Db 421 EISPEVCPAASTVTVSPAVFVSPASSAVLPAVSLVPLTASVTSPKASPTVSPAAAPPT 480

Qy 481 ASPANKOVSSLETADVEETGEGLTASGGDVMMRRRIATPEEVRLPLQHGWRREVRIK 540
Db 481 ASPANKOVSSLETADVEETGEGLTASGGDVMMRRRIATPEEVRLPLQHGWRREVRIK 540

Qy 541 KGSHRWQETWYVYGCGRMKQFPPVVKYLSRNLVHVSVRREHFSFSPMPVGDFFEEEDT 600
Db 541 KGSHRWQETWYVYGCGRMKQFPPVVKYLSRNLVHVSVRREHFSFSPMPVGDFFEEEDT 600

Qy 601 PEGLOWVLSABEIPSRIOAITGRGRPRNTEKAKTEVPKVGKRGRRPPKVKITELNKK 660
Db 601 PEGLOWVLSABEIPSRIOAITGRGRPRNTEKAKTEVPKVGKRGRRPPKVKITELNKK 660

Qy 661 TDNRPLKLEAQETLNEDKAKIAKSKKMKQKVGQEGCLTTIOQAARKKQKETSLSKH 720
Db 661 TDNRPLKLEAQETLNEDKAKIAKSKKMKQKVGQEGCLTTIOQAARKKQKETSLSKH 720

Qy 721 KEAKKSAEKEKGTQKLEKEVKREKKEKVKMKEEETVKAPACKADTLATQRL 780
Db 721 KEAKKSAEKEKGTQKLEKEVKREKKEKVKMKEEETVKAPACKADTLATQRL 780

Qy 781 EERQKQMILEMKKPTEDMCLTDHQLPDPFSRVPGTLPSGAFSDCLTIVFELHSGKV 840
Db 781 EERQKQMILEMKKPTEDMCLTDHQLPDPFSRVPGTLPSGAFSDCLTIVFELHSGKV 840

Qy 841 LGFDPKADVPKSLGVLQEGLLCQGDLSGEVODLLVRLKKAALHDPGFPSYCOSLKILGEKV 900
Db 841 LGFDPKADVPKSLGVLQEGLLCQGDLSGEVODLLVRLKKAALHDPGFPSYCOSLKILGEKV 900

Qy 901 SEIPLTRDNVSEILRCFLMAYGVXPALCDRLRTOPFOAQQPQKAAVLAPVPHELNGSTL 960
Db 901 SEIPLTRDNVSEILRCFLMAYGVXPALCDRLRTOPFOAQQPQKAAVLAPVPHELNGSTL 960

Qy 961 IINEIDKTLESMSYRNKWIIVEGRRLRLKTVLAKRTGRSEVENGRPEECLGRRRSSIM 1020
Db 961 IINEIDKTLESMSYRNKWIIVEGRRLRLKTVLAKRTGRSEVENGRPEECLGRRRSSIM 1020

Qy 1021 BETSGMEEREEESIAAVPGRRRGRDGEVDATASSIPELEHQIEKLSKROLFFFRKLLHS 1080
Db 1021 BETSGMEEREEESIAAVPGRRRGRDGEVDATASSIPELEHQIEKLSKROLFFFRKLLHS 1080

Qy 1081 SQMLRAVSLGQDRYRRRYWVLPYLAGIFVEGTEGNLVPVEVIKKEITDSLKVAHAASLNPA 1140
Db 1081 SQMLRAVSLGQDRYRRRYWVLPYLAGIFVEGTEGNLVPVEVIKKEITDSLKVAHAASLNPA 1140

Qy 1141 LFSMKMELAGSNTTASSPARARSPLTKPGFMQPRHFKSPVRGQDSBPQAOLOPBAQL 1200
Db 1141 LFSMKMELAGSNTTASSPARARSPLTKPGFMQPRHFKSPVRGQDSBPQAOLOPBAQL 1200

Qy 1201 HVPAPQPOLQLOLQSHKGFLEQSGSPLSLQSOHDLSSQAFSLWSLQTSQSHSLLSSSV 1260
Db 1201 HVPAPQPOLQLOLQSHKGFLEQSGSPLSLQSOHDLSSQAFSLWSLQTSQSHSLLSSSV 1260

Qy 1261 LTPDSSPGKLDPAPOPPEPEPEDEAESPDQLAFWNI SAQMPCNAAPTPLPVSSEDQP 1320
Db 1261 LTPDSSPGKLDPAPOPPEPEPEDEAESPDQLAFWNI SAQMPCNAAPTPLPVSSEDQP 1320

Qy 1321 TPSQOQLASKPMNRPSAANPCSPVQFSSTPLAGLAKRRAGDGPGEPMQSGTGLGQPKRR 1380
Db 1321 TPSQOQLASKPMNRPSAANPCSPVQFSSTPLAGLAKRRAGDGPGEPMQSGTGLGQPKRR 1380

Qy 1381 GRPSKPFKQMEQRYLTQLTAQVPVPMCMCGMWIIPDEMLDAMLKALHPRGIREKALHK 1440
Db 1381 GRPSKPFKQMEQRYLTQLTAQVPVPMCMCGMWIIPDEMLDAMLKALHPRGIREKALHK 1440

Qy 1441 HLNKHDFLOEVCILRPSADPIFEPRQLPAFQEGIMSPKEKTYETDLAVLQWVEELEQR 1500
Db 1441 HLNKHDFLOEVCILRPSADPIFEPRQLPAFQEGIMSPKEKTYETDLAVLQWVEELEQR 1500

Qy 1501 VIMSDLOIRGWTCPSPDSTREDLAYCEHLSQSDSDITWRGPGREGGLAPQKRTTNPLDLAV 1560
Db 1501 VIMSDLOIRGWTCPSPDSTREDLAYCEHLSQSDSDITWRGPGREGGLAPQKRTTNPLDLAV 1560

Qy 1561 MRLAALQONVKRRYLRRPLMPTHEVLEKALLSTPNCAPEGTTTEISYEITPRIRIWRQT 1620
Db 1561 MRLAALQONVKRRYLRRPLMPTHEVLEKALLSTPNCAPEGTTTEISYEITPRIRIWRQT 1620

Qy 1621 LQRCRAAHVCLCLGHLERSIAWEKSVNKVTCVCRKGDNDDEFLLDCDGRGCHYIYCHR 1680
Db 1621 LQRCRAAHVCLCLGHLERSIAWEKSVNKVTCVCRKGDNDDEFLLDCDGRGCHYIYCHR 1680

Qy 1681 PKMEAPEGWDFCTVCLAAQVVEGFTQKPGPKRQKRGKSGYSINLFSEGDRRRRVLLKG 1740
Db 1681 PKMEAPEGWDFCTVCLAAQVVEGFTQKPGPKRQKRGKSGYSINLFSEGDRRRRVLLKG 1740

Qy 1741 RESPAAGPRYSEERLSPSKRRRLSMRNHSDLTCEIILMWMESHDAAWPFXEPVNPRLV 1800
Db 1741 RESPAAGPRYSEERLSPSKRRRLSMRNHSDLTCEIILMWMESHDAAWPFXEPVNPRLV 1800

Qy 1801 SGYRRIIKNPMDFSTMERLLRGYTSSEFAADALLVFDNCQTFNEDDSEVGKAGHIMR 1860
Db 1801 SGYRRIIKNPMDFSTMERLLRGYTSSEFAADALLVFDNCQTFNEDDSEVGKAGHIMR 1860

Qy 1861 RPFESRWEETVQKQANL 1878
Db 1861 RPFESRWEETVQKQANL 1878
```

RESULT 2

US-10-376-537-13
; Sequence 13, Application US/10376537
; Publication No. US20030224405A1
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042001
; CURRENT APPLICATION NUMBER: US/10/376,537
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US/09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 1878
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(1878)
; OTHER INFORMATION: Xaa = Any Amino Acid

US-10-376-537-13

Query Match 99.9%; Score 9816; DB 15; Length 1878;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1878; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MEMEANEANDHNF	TGLPAPAAAGL	KPSPSGEGLY	NGSPMPPOQK	SLNGDVNYG	60				
Db	1	MEMEANEANDHNF	TGLPAPAAAGL	KPSPSGEGLY	NGSPMPPOQK	SLNGDVNYG	60				
Qy	61	LSTVSHTTT	SGILNSAPHSST	SHLHP	SVAYDCLWNT	SOYPSANPG	SNLKDPLLSQFS	120			
Db	61	LSTVSHTTT	SGILNSAPHSST	SHLHP	SVAYDCLWNT	SOYPSANPG	SNLKDPLLSQFS	120			
Qy	121	GQYPLNG	ILGGSRPSP	SHNTNL	RAGSKFWANG	THSPMGLN	FDSDQLYD	180			
Db	121	GQYPLNG	ILGGSRPSP	SHNTNL	RAGSKFWANG	THSPMGLN	FDSDQLYD	180			
Qy	181	EVCSGIHP	DEAAEKEMT	SVVAENG	TGLVCSLE	LEEXQPEL	KMCGYNG	SVPSVESHQEV	240		
Db	181	EVCSGIHP	DEAAEKEMT	SVVAENG	TGLVCSLE	LEEXQPEL	KMCGYNG	SVPSVESHQEV	240		
Qy	241	VLVPDPT	VSCLD	DPDPSH	LDPQLED	TPLGED	SLPEPNS	LAPVPVSG	GLYDID	300	
Db	241	VLVPDPT	VSCLD	DPDPSH	LDPQLED	TPLGED	SLPEPNS	LAPVPVSG	GLYDID	300	
Qy	301	KLPLXDS	SPVISAL	DCPSLNN	ATAFSL	LADD	SDTST	SIFASPT	SPVVLGES	VLQD	360
Db	301	KLPLXDS	SPVISAL	DCPSLNN	ATAFSL	LADD	SDTST	SIFASPT	SPVVLGES	VLQD	360
Qy	361	NGSDAEQ	EMETQSSD	FPPLTQ	PADQSS	TTQLHP	ATSPAV	SPPTSP	AVSLVSP	PAASP	420
Db	361	NGSDAEQ	EMETQSSD	FPPLTQ	PADQSS	TTQLHP	ATSPAV	SPPTSP	AVSLVSP	PAASP	420
Qy	421	EISPEV	CPAAST	VSPV	SVSPAS	SAVLPA	VSLE	VPLTAS	VTSPK	ASPVTS	480
Db	421	EISPEV	CPAAST	VSPV	SVSPAS	SAVLPA	VSLE	VPLTAS	VTSPK	ASPVTS	480
Qy	481	ASPANKD	VSSFL	ETAD	VEET	IGEL	TASG	SDVNR	RIATPEE	VRPLQ	540
Db	481	ASPANKD	VSSFL	ETAD	VEET	IGEL	TASG	SDVNR	RIATPEE	VRPLQ	540
Qy	541	KGSHRWQ	ETWYGP	CGKRM	KQFP	EV	IKYLS	RNLV	HSVR	REHFS	600

Db	541	KGSHRWQ	ETWYGP	CGKRM	KQFP	EV	IKYLS	RNLV	HSVR	REHFS	600
Qy	601	PEGLQW	QLSA	EEIP	SRIQ	AITG	KRGR	PNTE	KAKT	KEVP	660
Db	601	PEGLQW	QLSA	EEIP	SRIQ	AITG	KRGR	PNTE	KAKT	KEVP	660
Qy	661	TDNRP	LKLE	AOPT	LN	EEED	KAKI	AKSK	KMRQ	KVOR	720
Db	661	TDNRP	LKLE	AOPT	LN	EEED	KAKI	AKSK	KMRQ	KVOR	720
Qy	721	KEAKK	SAE	KEG	KGT	KQEK	LKEV	KREK	KVK	KWKE	780
Db	721	KEAKK	SAE	KEG	KGT	KQEK	LKEV	KREK	KVK	KWKE	780
Qy	781	EERQ	KQW	ILE	EMK	KPT	EDM	CLTD	HQPL	PDF	840
Db	781	EERQ	KQW	ILE	EMK	KPT	EDM	CLTD	HQPL	PDF	840
Qy	841	LGFP	PAK	DV	PS	LG	VL	QEG	SL	GEV	900
Db	841	LGFP	PAK	DV	PS	LG	VL	QEG	SL	GEV	900
Qy	901	SEI	PL	TR	DV	SEI	IL	RC	FL	MA	960
Db	901	SEI	PL	TR	DV	SEI	IL	RC	FL	MA	960
Qy	961	II	NEI	D	K	T	LE	S	M	S	1020
Db	961	II	NEI	D	K	T	LE	S	M	S	1020
Qy	1021	BET	S	G	M	B	E	E	E	E	1080
Db	1021	BET	S	G	M	B	E	E	E	E	1080
Qy	1081	SQ	M	L	R	A	V	S	L	G	1140
Db	1081	SQ	M	L	R	A	V	S	L	G	1140
Qy	1141	L	F	S	M	K	M	E	L	A	1200
Db	1141	L	F	S	M	K	M	E	L	A	1200
Qy	1201	H	V	P	A	Q	P	O	L	O	1260
Db	1201	H	V	P	A	Q	P	O	L	O	1260
Qy	1261	L	T	P	D	S	S	P	G	K	1320
Db	1261	L	T	P	D	S	S	P	G	K	1320
Qy	1321	T	P	S	P	Q	L	A	S	K	1380
Db	1321	T	P	S	P	Q	L	A	S	K	1380
Qy	1381	G	R	P	S	K	F	K	Q	E	1440
Db	1381	G	R	P	S	K	F	K	Q	E	1440
Qy	1441	H	L	N	K	H	R	D	F	L	1500
Db	1441	H	L	N	K	H	R	D	F	L	1500
Qy	1501	V	I	M	S	D	L	I	R	G	1560
Db	1501	V	I	M	S	D	L	I	R	G	1560
Qy	1561	M	R	L	A	E	Q	N	V	K	1620
Db	1561	M	R	L	A	E	Q	N	V	K	1620
Qy	1621	L	Q	R	C	S	A	H	V	C	1680
Db	1621	L	Q	R	C	S	A	H	V	C	1680

QY 1681 PKMEAVPBGDFCTVCLAQOVEGFTQKPPKRGKSGVSLNFSGDRRRVLLKG 1740
DB 1681 PKMEAVPBGDFCTVCLAQOVEGFTQKPPKRGKSGVSLNFSGDRRRVLLKG 1740
QY 1741 RESPAAGPRYSEERLSPSKRRRLSRNHHSDLTFCETIILMESHDAAMPFXEPVNPRLV 1800
DB 1741 RESPAAGPRYSEERLSPSKRRRLSRNHHSDLTFCETIILMESHDAAMPFXEPVNPRLV 1800
QY 1801 SGYRIIKNPMDFTSMRRLRGGYTSSSEFAADALLVFDNCQTFNEDDSEVGKAGHIMR 1860
DB 1801 SGYRIIKNPMDFTSMRRLRGGYTSSSEFAADALLVFDNCQTFNEDDSEVGKAGHIMR 1860
QY 1861 RFFESRWEIFYGKQKQANL 1878
DB 1861 RFFESRWEIFYGKQKQANL 1878
RESULT 3
US-10-702-148-13
; Sequence 13, Application US/10702148
; Publication No. US20040063145A1
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042002
; CURRENT APPLICATION NUMBER: US/10/702,148
; CURRENT FILING DATE: 2003-11-05
; PRIOR APPLICATION NUMBER: PRIOR APPLICATION NUMBER: US/09/839,479
; PRIOR FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JF98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 1878
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(1878)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-10-702-148-13
Query Match 99.9%; Score 9816; DB 15; Length 1878;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1878; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MEMEANEANDHFNFTGLPPAAPAGLKPSPSSGGLYTNNGSPMFPQOGKSLNGDVNNG 60
DB 1 MEMEANEANDHFNFTGLPPAAPAGLKPSPSSGGLYTNNGSPMFPQOGKSLNGDVNNG 60
QY 61 LSTVSHTTTSGILNSAPHSSSTSHLHPHSVAYDCLMNYSQYPSANPGSNLKDPPLLSQFS 120
DB 61 LSTVSHTTTSGILNSAPHSSSTSHLHPHSVAYDCLMNYSQYPSANPGSNLKDPPLLSQFS 120
QY 121 GGQYPLNGILGGRSQPSPSHNTNLRAGSQFWANGTHSPMGLNFDPSQELYDSPDQNF 180
DB 121 GGQYPLNGILGGRSQPSPSHNTNLRAGSQFWANGTHSPMGLNFDPSQELYDSPDQNF 180
QY 181 EVCSGIHDEAREKEMTSVVAENGTGLVCSLEEXQPELKWCCYNGSVPSVESLHOEVS 240
DB 181 EVCSGIHDEAREKEMTSVVAENGTGLVCSLEEXQPELKWCCYNGSVPSVESLHOEVS 240
QY 241 VLVPDPTVSCLDPSHLPDQLEDTPILSEDSLEFPNSLAPFVSGGLYGIDDTLMGAED 300
DB 241 VLVPDPTVSCLDPSHLPDQLEDTPILSEDSLEFPNSLAPFVSGGLYGIDDTLMGAED 300

QY 301 KLPLXDSPTVISALDCPSILNNATAFSLIADDSQTSSTIPASPTSPVLGESVLQNSFDLN 360
DB 301 KLPLXDSPTVISALDCPSILNNATAFSLIADDSQTSSTIPASPTSPVLGESVLQNSFDLN 360
QY 361 NGSDAEQEBMETQSSDPPPSLTQAPDQSSSTIQLHPATSPAVSPTTSPAVSLVSPAAASP 420
DB 361 NGSDAEQEBMETQSSDPPPSLTQAPDQSSSTIQLHPATSPAVSPTTSPAVSLVSPAAASP 420
QY 421 EISPEVCPAASTVVSFAVSVSPASSAVILPAVLEVLPLTASVTSPKASPTVSPAAAPPT 480
DB 421 EISPEVCPAASTVVSFAVSVSPASSAVILPAVLEVLPLTASVTSPKASPTVSPAAAPPT 480
QY 481 ASPANKDVSSPLETTADVVEITGEGLTASGSDGVMMRRRIATPEVRPLQHGWRREVRK 540
DB 481 ASPANKDVSSPLETTADVVEITGEGLTASGSDGVMMRRRIATPEVRPLQHGWRREVRK 540
QY 541 KGSHRWQGETWYGYPCGKRMKQFPEVIKYLRSNLVHSVRRHFSSFRMPVGDFFFEEDT 600
DB 541 KGSHRWQGETWYGYPCGKRMKQFPEVIKYLRSNLVHSVRRHFSSFRMPVGDFFFEEDT 600
QY 601 PEGLOWVOLSABEIPSRIOAITGKRGPRNTEKATKEVPKVKRGGRGPPKVKITELINK 660
DB 601 PEGLOWVOLSABEIPSRIOAITGKRGPRNTEKATKEVPKVKRGGRGPPKVKITELINK 660
QY 661 TDNRPLAKLFAOETLNEEDKAKIAKSKKWKQVORGECLTTIOGOARNKEKOETKSLKH 720
DB 661 TDNRPLAKLFAOETLNEEDKAKIAKSKKWKQVORGECLTTIOGOARNKEKOETKSLKH 720
QY 721 KEAKKKSXAEEKGKTQOEKLKEVKREKEKVKMKEKEEVTAKPACADKTLATORRL 780
DB 721 KEAKKKSXAEEKGKTQOEKLKEVKREKEKVKMKEKEEVTAKPACADKTLATORRL 780
QY 781 EERQKQMILEEMKKPTEDMCLTDHQLPDPFSRVPGLTLPFGAFSDCLTIVEFLHSPGKV 840
DB 781 EERQKQMILEEMKKPTEDMCLTDHQLPDPFSRVPGLTLPFGAFSDCLTIVEFLHSPGKV 840
QY 841 LGFDPKADVPSLGVLOEGLICQSDSLGEVODLLVRLKAAALHDPFGFSYCSOLKILGKV 900
DB 841 LGFDPKADVPSLGVLOEGLICQSDSLGEVODLLVRLKAAALHDPFGFSYCSOLKILGKV 900
QY 901 SEIPLTRDNVSEILRCLFMAYGVXALCDRLRTOFPOAQPPQKAAVLAFFVHNLNGSTL 960
DB 901 SEIPLTRDNVSEILRCLFMAYGVXALCDRLRTOFPOAQPPQKAAVLAFFVHNLNGSTL 960
QY 961 IINEIDKTLSEMSYRKNKWIVEGRRLRLKTVLAKRTGRSEVENGRPEECILGRRSSRIM 1020
DB 961 IINEIDKTLSEMSYRKNKWIVEGRRLRLKTVLAKRTGRSEVENGRPEECILGRRSSRIM 1020
QY 1021 EETSGMEEEEEEESIAAVPGRGRGRDGEVDATASSIPELEROIEKLSKROLFFRKLLHS 1080
DB 1021 EETSGMEEEEEEESIAAVPGRGRGRDGEVDATASSIPELEROIEKLSKROLFFRKLLHS 1080
QY 1081 SQMLRAVSLGQDRYRRRYWVLPYLAGIFVGEQTEGNLVPEEVIKKEKTDLSKVAHAASLNPA 1140
DB 1081 SQMLRAVSLGQDRYRRRYWVLPYLAGIFVGEQTEGNLVPEEVIKKEKTDLSKVAHAASLNPA 1140
QY 1141 LFSMKMELAGSNNTTASSPARARSRLTKTGFQMPQPHFKGSPVRQDSEQQAQLQPEAQL 1200
DB 1141 LFSMKMELAGSNNTTASSPARARSRLTKTGFQMPQPHFKGSPVRQDSEQQAQLQPEAQL 1200
QY 1201 HVPAPQPOLOLOLQSHKGFLEQEGSPISLQSOQHDLSQSAFLSWLSQTSQSHSLLSSSV 1260
DB 1201 HVPAPQPOLOLOLQSHKGFLEQEGSPISLQSOQHDLSQSAFLSWLSQTSQSHSLLSSSV 1260
QY 1261 LTPDSSFGKLDPAQSPPEPEPEDEAESSDLOQAFWNIQAQPCNAAPTPLAVSEDOQ 1320
DB 1261 LTPDSSFGKLDPAQSPPEPEPEDEAESSDLOQAFWNIQAQPCNAAPTPLAVSEDOQ 1320
QY 1321 TSPSQQLASSKPMNRPSAANPCSPVQFSSTPLAGLAKRRAGDPGEMPSQSTGLGQPKRR 1380
DB 1321 TSPSQQLASSKPMNRPSAANPCSPVQFSSTPLAGLAKRRAGDPGEMPSQSTGLGQPKRR 1380

```
Qy 1381 GRPSPKFKQMEORYLTQITQAPVPPCMSCGHWIIPDEMLDAMLKALHPRGIREKALHK 1440
Db 1381 GRPSPKFKQMEORYLTQITQAPVPPCMSCGHWIIPDEMLDAMLKALHPRGIREKALHK 1440
Qy 1441 HLNKHRDFLOEVCVLRPSADPIFEPRQLPAFOEGIMSWSPKKTETDLDLAVLQWVEELBQR 1500
Db 1441 HLNKHRDFLOEVCVLRPSADPIFEPRQLPAFOEGIMSWSPKKTETDLDLAVLQWVEELBQR 1500
Qy 1501 VIMSDQIRGWTCPSPDSTREDLAYCEHLSDSQEDITWRGPGREGLAPOKRTTNPLDLAV 1560
Db 1501 VIMSDQIRGWTCPSPDSTREDLAYCEHLSDSQEDITWRGPGREGLAPOKRTTNPLDLAV 1560
Qy 1561 MRLAALQNVKRYLRRELPWTHEVYLEKALLSTPNGAPEGTTTISEIETPRIRINROT 1620
Db 1561 MRLAALQNVKRYLRRELPWTHEVYLEKALLSTPNGAPEGTTTISEIETPRIRINROT 1620
Qy 1621 LORCSAAHVCLCLGHLSRTAWKSVNKTCLVCRKGDNDDEFLLCDGCRGCHYCHR 1680
Db 1621 LORCSAAHVCLCLGHLSRTAWKSVNKTCLVCRKGDNDDEFLLCDGCRGCHYCHR 1680
Qy 1681 PKMEAVPBGDFWCTVCLAQOVEGFTQKPGPKRQKRGKSGVSLNFSFGDGRRRVLLKG 1740
Db 1681 PKMEAVPBGDFWCTVCLAQOVEGFTQKPGPKRQKRGKSGVSLNFSFGDGRRRVLLKG 1740
Qy 1741 RESPAAGPRYSEERLSPSKRRRLSRNHHSDLTFCCEIILMEMESHDAAWPXPVPVNPRLV 1800
Db 1741 RESPAAGPRYSEERLSPSKRRRLSRNHHSDLTFCCEIILMEMESHDAAWPXPVPVNPRLV 1800
Qy 1801 SGYRIILKNPWFDMFMRERLARGVTSSEEFPAADALLVFDNCQTFNEDDSEVGKAGHIMR 1860
Db 1801 SGYRIILKNPWFDMFMRERLARGVTSSEEFPAADALLVFDNCQTFNEDDSEVGKAGHIMR 1860
Qy 1861 RFFESRWEFFYQGGQANL 1878
Db 1861 RFFESRWEFFYQGGQANL 1878

RESULT 4
US-09-839-479-70
; Sequence 70, Application US/09839479
; Publication No. US20020039779A1
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042002
; CURRENT APPLICATION NUMBER: US/09/839,479
; CURRENT FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 70
; LENGTH: 1876
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-839-479-70

Query Match 98.8%; Score 9712; DB 9; Length 1876;
Best Local Similarity 99.1%; Pred. No. 0;
Matches 1861; Conservative 1; Mismatches 14; Indels 2; Gaps 2;

Qy 1 MEMEANEANDHFNFTGLPPAAPASGLKPSGSGGLYTNNGSPMNPPOQKKSINGDVNNG 60
Db 1 MEMEANEANDHFNFTGLPPAAPASGLKPSGSGGLYTNNGSPMNPPOQKKSINGDVNNG 60
Qy 61 LSTVSHTTTSGILNSAPHSSSTSHLHPSVAYDCLWNYSQYPSANPGSNLKDPPLLSQFS 120
Db 61 LSTVSHTTTSGILNSAPHSSSTSHLHPSVAYDCLWNYSQYPSANPGSNLKDPPLLSQFS 120
```

```
Db 61 LSTVSHTTTSGILNSAPHSSSTSHLHPSVAYDCLWNYSQYPSANPGSNLKDPPLLSQFS 120
Qy 121 GGQYPLNGILGSGRQSPSPSHNTNLRAGSQKFWANGTHSPMGLNFDSDQELYDSDPDQNF 180
Db 121 GGQYPLNGILGSGRQSPSPSHNTNLRAGSQKFWANGTHSPMGLNFDSDQELYDSDPDQNF 180
Qy 181 EVCSGIHPDEAAEKEMTSVAENGTGLVCSLESEEXQPELXKMGYNGSVPSVESLHQEVS 240
Db 181 EVCSGIHPDEAAEKEMTSVAENGTGLVCSLESEEXQPELXKMGYNGSVPSVESLHQEVS 240
Qy 241 VLVPDPTVSCLDLDDSHLPDQLEDPTILSEDSLEPNSLAPPEPVSGGLXGIDDTLGMGAED 300
Db 241 VLVPDPTVSCLDLDDSHLPDQLEDPTILSEDSLEPNSLAPPEPVSGGLXGIDDTLGMGAED 300
Qy 301 KLPLXDSPVISALDCPSLNNATAFSLIADDSQTSISIPASPTSPVLGESVLQDQNSFDLN 360
Db 301 KLPLXDSPVISALDCPSLNNATAFSLIADDSQTSISIPASPTSPVLGESVLQDQNSFDLN 360
Qy 361 NGSDAEOEEMETQSDPPPSLTQAPDOSSTIQLHPATSPAVSTTSPAVSLVSPAAASP 420
Db 361 NGSDAEOEEMETQSDPPPSLTQAPDOSSTIQLHPATSPAVSTTSPAVSLVSPAAASP 420
Qy 421 EISPEVCPAASTVVS PAVFSVWSPASSAVLPAVLEVLPTASVTS PKASPVTSPAAAPPT 480
Db 421 EISPEVCPAASTVVS PAVFSVWSPASSAVLPAVLEVLPTASVTS PKASPVTSPAAAPPT 480
Qy 481 ASPANKDVSSFLETTADVVEITGEGLTASGSGDVMMRRRIATPEEVRLPLQHGWRREVRIK 540
Db 481 ASPANKDVSSFLETTADVVEITGEGLTASGSGDVMMRRRIATPEEVRLPLQHGWRREVRIK 540
Qy 541 KGSHRWQGETWYYPGCGKRMKQFPPEVIKYLNRNLVHVSRRRHFSFSPMPVGDFFEEEDT 600
Db 541 KGSHRWQGETWYYPGCGKRMKQFPPEVIKYLNRNLVHVSRRRHFSFSPMPVGDFFEEEDT 600
Qy 601 PEGLOWVLSABEIPSRIOAITGKGRPRNTEKAKTKEVPKVKRGGRGPPVKVKTTELLNK 660
Db 601 PEGLOWVLSABEIPSRIOAITGKGRPRNTEKAKTKEVPKVKRGGRGPPVKVKTTELLNK 660
Qy 661 TDNRPLKKLEAQETLNEEDKAKIAKSKKQMRQKQVORGECLTTIIOQARNKKEQETKSLKH 720
Db 661 TDNRPLKKLEAQETLNEEDKAKIAKSKKQMRQKQVORGECLTTIIOQARNKKEQETKSLKH 720
Qy 721 KEAKKSKAEKEKTKQEKUKEVKREKKEKVKQKKEEVTAKKAPACKADKTATQRL 780
Db 721 KEAKKSKAEKEKTKQEKUKEVKREKKEKVK- KEKEEVTAKKAPACKADKTATQRL 779
Qy 781 EEROKQOMILEEMKPTEDMCLTDHQLPDFSRVPGLTLP SGAFSDCLTIIVEFLHSFGKV 840
Db 781 EEROKQOMILEEMKPTEDMCLTDHQLPDFSRVPGLTLP SGAFSDCLTIIVEFLHSFGKV 839
Qy 841 LGFDPKDVPSLGLVQLQEGLLCOGDSLGEVQDILLVRLKKAALHDPGFPSYCQSLKILGEKV 900
Db 841 LGFDPKDVPSLGLVQLQEGLLCOGDSLGEVQDILLVRLKKAALHDPGFPSYCQSLKILGEKV 899
Qy 901 SEIPLTRDNVSEILRCLFMAYGVXPALCDRLRTQFQAQPPQKAAVLAPVHIELNGSTL 960
Db 901 SEIPLTRDNVSEILRCLFMAYGVXPALCDRLRTQFQAQPPQKAAVLAPVHIELNGSTL 959
Qy 961 IINEIDKTLESMSYRNKNWIVEGLRLKTVLAKRTGRSEVENGRPEECIGRRSSRIM 1020
Db 961 IINEIDKTLESMSYRNKNWIVEGLRLKTVLAKRTGRSEVENGRPEECIGRRSSRIM 1019
Qy 1021 EETSGMEEEESIAAVPGRGRRDGEVDATASSIPELERQIEKLSKROLFFRKKLLHS 1080
Db 1021 EETSGMEEEESIAAVPGRGRRDGEVDATASSIPELERQIEKLSKROLFFRKKLLHS 1079
Qy 1081 SQMLRAVSLGQDRYRRRYVWLPYLAGIFVEGTEGNLVPEEVIKKETDSLKVAHAASLNP 1140
Db 1081 SQMLRAVSLGQDRYRRRYVWLPYLAGIFVEGTEGNLVPEEVIKKETDSLKVAHAASLNP 1139
Qy 1141 LFSMKMELAGSNNTTASSPARARSPLKTKPGFMQPREFKSPVRGQDSQPOAQLOPEAQL 1200
Db 1141 LFSMKMELAGSNNTTASSPARARSPLKTKPGFMQPREFKSPVRGQDSQPOAQLOPEAQL 1199
```


Qy 1201 HVPAQPOLOLOQSHKGFLEQBGSPLSLQSOHDLSSQAFSLWLSQTOSSHSLSSV 1260
Db 1200 HVPAQPOLOLOQSHKGFLEQBGSPLSLQSOHDLSSQAFSLWLSQTO-HSSLSSV 1258
Qy 1261 LTPDSSPGKLDPAQSPQPEPEDEASSPDLOAFWFNISAQMPCNAAPTTPPLAVSEDQP 1320
Db 1259 LTPDSSPGKLDPAQSPQPEPEDEASSPDLOAFWFNISAQMPCNAAPTTPPLAVSEDQP 1318
Qy 1321 TPSQQLASSKPMRPSAANPCSPVQSSPTPLAGLAPKRRAGDCEMPQSPGTGLGQPKRR 1380
Db 1319 TPSQQLASSKPMRPSAANPCSPVQSSPTPLAGLAPKRRAGDCEMPQSPGTGLGQPKRR 1378
Qy 1381 GRPSKFFKQMEQRYLTQTAOPVPEMCSSGWWIPDPEMLDAMLKALHPRGIREKALHK 1440
Db 1379 GRPSKFFKQMEQRYLTQTAOPVPEMCSSGWWIPDPEMLDAMLKALHPRGIREKALHK 1438
Qy 1441 HLNKHRDFLOEVCILRPSADPIFEPRQLPAFOEGIMSPKETYETDLAVLQWVEELEQR 1500
Db 1439 HLNKHRDFLOEVCILRPSADPIFEPRQLPAFOEGIMSPKETYETDLAVLQWVEELEQR 1498
Qy 1501 VIMSDQIRGWTCPSPDSTREDLAYCEHLSDSQSDITWRGPGREGLAPQKRTTNPLDLAV 1560
Db 1499 VIMSDQIRGWTCPSPDSTREDLAYCEHLSDSQSDITWRGPGREGLAPQKRTTNPLDLAV 1558
Qy 1561 MRLAALSONVKRRYLREPLWPTHEVLEKALLSTPNGAPEGTTTTEISYEITPRIRIWRQT 1620
Db 1559 MRLAALSONVKRRYLREPLWPTHEVLEKALLSTPNGAPEGTTTTEISYEITPRIRIWRQT 1618
Qy 1621 LQRCRAAHVCLCLGHLESLIAWKSVMKVTCVCRKGDNDDEFLLCDGCDRGCHYICHR 1680
Db 1619 LQRCRAAHVCLCLGHLESLIAWKSVMKVTCVCRKGDNDDEFLLCDGCDRGCHYICHR 1678
Qy 1681 PKMAVPEGDWFTCVCLAQVVEGFTQKPGPKGQKRGKSGYSNLFSEGDRRRVLLKG 1740
Db 1679 PKMAVPEGDWFTCVCLAQVVEGFTQKPGPKGQKRGKSGYSNLFSEGDRRRVLLKG 1738
Qy 1741 RESPAAGPRYSEERLSPSKRRRLSMRNHSDLTTCCELLMEMESHDAWPEXPVNPRLV 1800
Db 1739 RESPAAGPRYSEERLSPSKRRRLSMRNHSDLTTCCELLMEMESHDAWPEXPVNPRLV 1798
Qy 1801 SGYRRIKNPDMFSTMERLLRGYTSSEEPADALLVFNOCQFNEDDSSVGKAGHMR 1860
Db 1799 SGYRRIKNPDMFSTMERLLRGYTSSEEPADALLVFNOCQFNEDDSSVGKAGHMR 1858
Qy 1861 RFFESRWEFFYQKQANL 1878
Db 1859 RFFESRWEFFYQKQANL 1876

RESULT 5

US-10-376-537-71
; Sequence 71, Application US/10376537
; Publication No. US20030224405A1
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042001
; CURRENT APPLICATION NUMBER: US/10/376,537
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US/09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 71
; LENGTH: 1876
; TYPE: PRT

i ORGANISM: Homo sapiens
US-10-376-537-71
Query Match 98.8%; Score 9712; DB 15; Length 1876;
Best Local Similarity 99.1%; Pred. No. 0;
Matches 1861; Conservative 1; Mismatches 14; Indels 2; Gaps 2;
Qy 1 MEMEANANDHFNFTGLPPAPAAAGLKPSPSSGEGLYTNGSPMNFPOQGSKSLGDNVNVG 60
Db 1 MEMEANANDHFNFTGLPPAPAAAGLKPSPSSGEGLYTNGSPMNFPOQGSKSLGDNVNVG 60
Qy 61 LSTVSHTTTSGILNSAPHSSSTSHLHHPSVAYDCLWNYSQYPSANPGSNLKDPELLSQFS 120
Db 61 LSTVSHTTTSGILNSAPHSSSTSHLHHPSVAYDCLWNYSQYPSANPGSNLKDPELLSQFS 120
Qy 121 GGQYPLNGILGGRQSPSPSHNTNLRAGSKFWANGTHSPMGLNFDSELYDSDPDQNF 180
Db 121 GGQYPLNGILGGRQSPSPSHNTNLRAGSKFWANGTHSPMGLNFDSELYDSDPDQNF 180
Qy 181 EVCSGTHPDEAAEKEMTSVVAENGTGLVCSLEEEQPELKMCGYNGSVPSVSLHGEVS 240
Db 181 EVCSGTHPDEAAEKEMTSVVAENGTGLVCSLEEEQPELKMCGYNGSVPSVSLHGEVS 240
Qy 241 VLVDPDPTVSCLDPPSHLPDQLEDTPILSEDSLEPPFNSLAPEPVSGGLYGIDDTLMGAED 300
Db 241 VLVDPDPTVSCLDPPSHLPDQLEDTPILSEDSLEPPFNSLAPEPVSGGLYGIDDTLMGAED 300
Qy 301 KLPLXGSPVTSALDCPSLNNATAPSLAADDSTSTSFASPTSPPVLGESVLQNSFDLN 360
Db 301 KLPLEDSPVTSALDCPSLNNATAPSLAADDSTSTSFASPTSPPVLGESVLQNSFDLN 360
Qy 361 NGSDAEQEMETOSSDPPLSLTOPADPOSTIQLHPATSPAVSPTTSPAVSLVVSPPASP 420
Db 361 NGSDAEQEMETOSSDPPLSLTOPADPOSTIQLHPATSPAVSPTTSPAVSLVVSPPASP 420
Qy 421 EISPEVCPAASTVVSFAVFWSPASSAVLPVLSLEVPLTASVTSPKASPTVSPAAAFPT 480
Db 421 EISPEVCPAASTVVSFAVFWSPASSAVLPVLSLEVPLTASVTSPKASPTVSPAAAFPT 480
Qy 481 ASPANKDVVSFLETTADVEBITGEGLTASGSDVMRRRIATPBEVRLPLOGHWRREVRK 540
Db 481 ASPANKDVVSFLETTADVEBITGEGLTASGSDVMRRRIATPBEVRLPLOGHWRREVRK 540
Qy 541 KGSRWQGETWYCGPCGRMKQFPEVIKYLSRNLVHSVRREHFSFSPMPVGDFFERD 600
Db 541 KGSRWQGETWYCGPCGRMKQFPEVIKYLSRNLVHSVRREHFSFSPMPVGDFFERD 600
Qy 601 PEGLOWQLSABEIPSRIOAITGKRGPRNTEKAKTKEVPKVRGRGRPKVKITELLNK 660
Db 601 PEGLOWQLSABEIPSRIOAITGKRGPRNTEKAKTKEVPKVRGRGRPKVKITELLNK 660
Qy 661 TDNRPLKLEAQETLNEEDKAKIAKSKKVRQKVQVQGECLTTIQQOARNKRKQSTKSLKH 720
Db 661 TDNRPLKLEAQETLNEEDKAKIAKSKKVRQKVQVQGECLTTIQQOARNKRKQSTKSLKH 720
Qy 721 KEAKKSAEKEGKTKQEKLEKEVKKEKVKQKEEVTAKKAPACKADKTATORRL 780
Db 721 KEAKKSAEKEGKTKQEKLEKEVKKEKVKQKEEVTAKKAPACKADKTATORRL 779
Qy 781 EERQKQOMILEEMKPTEDMCLTDHOPDPFSRVPGTLTPSGAFSDCLTTVEFLHSPGKV 840
Db 780 EERQKQOMILEEMKPTEDMCLTDHOPDPFSRVPGTLTPSGAFSDCLTTVEFLHSPGKV 839
Qy 841 LGFDPADKVPVSLVGLQEGILCQSDSLGEVODLLVRLKAAALHDPFPFSYQCSLKILEKV 900
Db 840 LGFDPADKVPVSLVGLQEGILCQSDSLGEVODLLVRLKAAALHDPFPFSYQCSLKILEKV 899
Qy 901 SEIPLTRDNVSEILRCFLMAYGVXPCALCDRLRTPFOAQPOQKAAVLAPEVHELNGSTL 960
Db 900 SEIPLTRDNVSEILRCFLMAYGVXPCALCDRLRTPFOAQPOQKAAVLAPEVHELNGSTL 959
Qy 961 IINEIDKTLSSMSYRKNKWIVEGRRLRLKTLAKRTGRSEVENMRPEECILGRSSSRIM 1020

Db 960 IINEIDKTLSESSYRKKNWIVGRLRLKTVLAKRGRSEVEMGRPECEIGRRSSRIM 1019
Qy 1021 EETSGMEEBEESIAAIPGRRGRDGEVDATASSIPELERQIEKLSKRQFFFRKKLIHS 1080
Db 1020 EETSGMEEBEESIAAIPGRRGRDGEVDATASSIPELERQIEKLSKRQFFFRKKLIHS 1079
Qy 1081 SOMLRAVSLGQDRYRRRWVLPYLAGIVFEGTEGNLVPVEVVKETDSLKVAAHASLNP 1140
Db 1080 SOMLRAVSLGQDRYRRRWVLPYLAGIVFEGTEGNLVPVEVVKETDSLKVAAHASLNP 1139
Qy 1141 LFSMKMELAGNNTTASSPARARSRLKTKPGFMOPRHFKSPVRGDSQPOQAQLOPEAQL 1200
Db 1140 LFSMKMELAGNNTTASSPARARSRLKTKPGFMOPRHFKSPVRGDSQPOQAQLOPEAQL 1199
Qy 1201 HVPAPQPOLQLOLQSHKGFLEQEGSPLSLQSGQHDLSQSFLSLSQTSQSHSSLSSSV 1260
Db 1200 HVPAPQPOLQLOLQSHKGFLEQEGSPLSLQSGQHDLSQSFLSLSQTSQSHSSLSSSV 1258
Qy 1261 LTPDSSPGKLDPAQSPQPEEPEDBAESSPDLOAFWFNISQMPNCNAAPTPLAVSEDP 1320
Db 1259 LTPDSSPGKLDPAQSPQPEEPEDBAESSPDLOAFWFNISQMPNCNAAPTPLAVSEDP 1318
Qy 1321 TSPSQOLASSKPMNRPSAANPCSPVQFSSPTPLAGLAPKRRAGDPCGMPQSGTGLGQPKRR 1380
Db 1319 TSPSQOLASSKPMNRPSAANPCSPVQFSSPTPLAGLAPKRRAGDPCGMPQSGTGLGQPKRR 1378
Qy 1381 GRPSPKFFKQMEQRYLTQTAQVPPENCSCGWWIIPDPEMLDAMLKALHPRGIREKALHK 1440
Db 1379 GRPSPKFFKQMEQRYLTQTAQVPPENCSCGWWIIPDPEMLDAMLKALHPRGIREKALHK 1438
Qy 1441 HLNKHRDPLQEVCLRPSADPIPEPRQLPAFQEGIMSWSPKKTETDIAVLQWVEELEQR 1500
Db 1439 HLNKHRDPLQEVCLRPSADPIPEPRQLPAFQEGIMSWSPKKTETDIAVLQWVEELEQR 1498
Qy 1501 VIMSDQIRGWTCPSPDSTRDLAYCEHLSQSDIITWRGPGREGLAPQRTTNPDLAV 1560
Db 1499 VIMSDQIRGWTCPSPDSTRDLAYCEHLSQSDIITWRGPGREGLAPQRTTNPDLAV 1558
Qy 1561 MRLAALQNVKRRYLREPLWPTHEVLEKALLSTPNGAPEGTTTIEISYEITPRIRIWRQT 1620
Db 1559 MRLAALQNVKRRYLREPLWPTHEVLEKALLSTPNGAPEGTTTIEISYEITPRIRIWRQT 1618
Qy 1621 LQRCRAAHVCLIGHLERSIAWEKSNVKTCLVCRKGDNDDEFLLCDGCDRGCHYCHR 1680
Db 1619 LQRCRAAHVCLIGHLERSIAWEKSNVKTCLVCRKGDNDDEFLLCDGCDRGCHYCHR 1678
Qy 1681 PKMEAVPGDFWCTVCLAQOQVEGTEQKPGKRGOKKSGYSLNFSFGDGRRRVLLKG 1740
Db 1679 PKMEAVPGDFWCTVCLAQOQVEGTEQKPGKRGOKKSGYSLNFSFGDGRRRVLLKG 1738
Qy 1741 RESPAAGPRYSEERLSPSKRRRLSMRNHSDLTFCCEIILMEMESHDAAPFEPVNPRLV 1800
Db 1739 RESPAAGPRYSEERLSPSKRRPLSMRNHSDLTFCCEIILMEMESHDAAPFEPVNPRLV 1798
Qy 1801 SGYRRIIKNPMDFSTMRERLLRGYTSSEFAADALLVFDNCQTFNEDDSEVGKAGHIMR 1860
Db 1799 SGYRRIIKNPMDFSTMRERLLRGYTSSEFAADALLVFDNCQTFNEDDSEVGKAGHIMR 1858
Qy 1861 RFFFSRWEEFYQKQANL 1878
Db 1859 RFFFSRWEEFYQKQANL 1876

RESULT 6
US-10-702-148-70
; Sequence 70, Application US/10702148
; Publication No. US20040063145A1
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042002
; CURRENT APPLICATION NUMBER: US/10/702,148
; CURRENT FILING DATE: 2003-11-05

; PRIOR APPLICATION NUMBER: PRIOR APPLICATION NUMBER: US/09/839,479
; PRIOR FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 70
; LENGTH: 1876
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-702-148-70

Query Match 98.8%; Score 9712; DB 15; Length 1876;
Best Local Similarity 99.1%; Pred. No. 0;
Matches 1861; Conservative 1; Mismatches 14; Indels 2; Gaps 2;

Qy 1 MEMEANANDHFNFTGLPPAPAAAGSLKPSGSSGGLYTNNGSPMNFPOQKSLNGDVNNG 60
Db 1 MEMEANANDHFNFTGLPPAPAAAGSLKPSGSSGGLYTNNGSPMNFPOQKSLNGDVNNG 60

Qy 61 LSTVSHTTTSGILNSAPHSSSTSHLHHPVAVAYDCLWNYSOYPSANPGSNLKDPLLSQFS 120
Db 61 LSTVSHTTTSGILNSAPHSSSTSHLHHPVAVAYDCLWNYSOYPSANPGSNLKDPLLSQFS 120

Qy 121 GGQYPLNGILGGRSQSPSSHNTNLRAGSQKFHWANGTHSPMGLNFDQSELVDSPPDQNF 180
Db 121 GGQYPLNGILGGRSQSPSSHNTNLRAGSQKFHWANGTHSPMGLNFDQSELVDSPPDQNF 180

Qy 181 EVCSGIHPDEAAEKEMTSVVAENGTGLVCSLEEXOPELKMCCGVNGSVPSVESHQSVS 240
Db 181 EVCSGIHPDEAAEKEMTSVVAENGTGLVCSLEEXOPELKMCCGVNGSVPSVESHQSVS 240

Qy 241 VLVPDPTVSCLDLDDPSHLPDQLEDPTILSEDSLEPFNSLAPPEPVSGGLYGIIDDTELMAED 300
Db 241 VLVPDPTVSCLDLDDPSHLPDQLEDPTILSEDSLEPFNSLAPPEPVSGGLYGIIDDTELMAED 300

Qy 301 KLPLXDSFVIALDCPSLNNATAPSLDADDSTSTSFASPTSPVVLGESVLQNSFDLN 360
Db 301 KLPLEDSFVIALDCPSLNNATAPSLDADDSTSTSFASPTSPVVLGESVLQNSFDLN 360

Qy 361 NGSDAEQEEMETQSDPFPPLTQAPQDSSTIQLHPATSPAVSPTTSPAVSLVSPASP 420
Db 361 NGSDAEQEEMETQSDPFPPLTQAPQDSSTIQLHPATSPAVSPTTSPAVSLVSPASP 420

Qy 421 EISPEVCPAASTVVSFAVSVSPASSAVLPAVSLVPLTASVTSKASPVTSAAAAPT 480
Db 421 EISPEVCPAASTVVSFAVSVSPASSAVLPAVSLVPLTASVTSKASPVTSAAAAPT 480

Qy 481 ASPANKOVSSFLETTADVVEITGEGLTASGSDVMRRRIATPVEVRLPQHGWRREVRK 540
Db 481 ASPANKOVSSFLETTADVVEITGEGLTASGSDVMRRRIATPVEVRLPQHGWRREVRK 540

Qy 541 KGSHRWQGETWYQPCGKRMKQFPEVIKYLNRNLVHVSRRHFSRPMVGDPEERDT 600
Db 541 KGSHRWQGETWYQPCGKRMKQFPEVIKYLNRNLVHVSRRHFSRPMVGDPEERDT 600

Qy 601 PEGLOWQLSAEETPSRIQAITGKGRPRNTEKAKTKEVPKVKGRGRPPKVKITELINK 660
Db 601 PEGLOWQLSAEETPSRIQAITGKGRPRNTEKAKTKEVPKVKGRGRPPKVKITELINK 660

Qy 661 TDNRPLKLEAQETLNEEDKAKIAKSKKMRQKVGORGECLTTIOQOARNKKEQTKSLKH 720
Db 661 TDNRPLKLEAQETLNEEDKAKIAKSKKMRQKVGORGECLTTIOQOARNKKEQTKSLKH 720

Qy 721 KEAKKSKAAEKEKTKQEKLEKVKREKKEKVKQKEEVTKAKPACKADKTATQRL 780
Db 721 KEAKKSKAAEKEKTKQEKLEKVKREKKEKVK- KEKEEVTKAKPACKADKTATQRL 779

QY 601 PEGLOWVLSABEISRSIOAITGKGRPRNTEKAKTKEVPKVRGRGPPKVKITELLNK 660
DB 598 PEGLOWVLSABEISRSIOAITGKGRPRNTEKAKTKEVPKVRGRGPPKVKITELLNK 657
QY 661 TDNRPLKLEAQETLNEDKAKIAKSKKMRQKVGCECLTTIOQOARKKROETKSLKH 720
DB 658 TDNRPLKLEAQETLNEDKAKIAKSKKMRQKVGCECQTTIOQOARKKROETKSLKQ 717
QY 721 KEAKKKSXAEBKGTKEKLEKVKREKKEKVKMKEEVTAKKAPACKADTKLATQRRLL 780
DB 718 KEAKKKS--KKEKGTKEKLEKVKREKKEKVKMKEEVTAKKAPACKADTKLATQRRLL 775
QY 781 EERQOQMLBEMKKPTEDMCLTDHOPLPDSRPVGLTLPGSFSDCLTIVFPLHSFGKV 840
DB 776 EERQOQMLBEMKKPTEDMCLTDHOPLPDSRPVGLTLPGSFSDCLTIVFPLHSFGKV 835
QY 841 LGFDPKDVPSLGVLOEGLLQCGDSLGEVDLLVRLKKAALHDPFPSPCOSLKLIGKV 900
DB 836 LGFDPKDVPSLGVLOEGLLQCGDSLGEVDLLVRLKKAALHDPFPSPCOSLKLIGKV 895
QY 901 SEIPLTRDNVSEILRCFLMAYGVXPALCDRLRTPFOAQQKAAVLAFVPHELNGSTL 960
DB 896 SEIPLTRDNVSEILRCFLMAYGVXPALCDRLRTPFOAQQKAAVLAFVPHELNGSTL 955
QY 961 IINEIDKTLSESSYRKWKWIVGRLRLKTVLAKRTGRSEVEMGRPEECGLRRSSRIM 1020
DB 956 IINEIDKTLSESSYRKWKWIVGRLRLKTVLAKRTGRSEVEMGRPEECGLRRSSRIM 1015
QY 1021 EETSMEEEEEESTAAVPGRRGRDGEVDATASSIPELERQIEKLSKROLFFRKLLHS 1080
DB 1016 EETSMEEEEEESTAAVPGRRGRDGEVDATASSIPELERQIEKLSKROLFFRKLLHS 1075
QY 1081 SOMRAVSLGQDRYRRYRWLPYLAGIVFEGTEGNLVPEEVIKKTETDSLKVAHAASLNPA 1140
DB 1076 SOMRAVSLGQDRYRRYRWLPYLAGIVFEGTEGNLVPEEVIKKTETDSLKVAHAASLNPA 1135
QY 1141 LFSMKMELAGNNTTASSPARARSRLPKYKPGMQPRHPKSPVRGQDSQOQAQLQPEAQL 1200
DB 1136 LFSMKMELAGNNTTASSPARARSRLPKYKPGMQPRHPKSPVRGQDSQOQAQLQPEAQL 1195
QY 1201 HVPQOPQLOLOLQSHKGFLEOESPLSLGSOHDLSQSAFLSWLSQTSQSHSLSSSV 1260
DB 1196 HAPAQOPQLOLOLQSHKGFLEOESPLSLGSOHDLSQSAFLSWLSQTSQSHSLSSSV 1255
QY 1261 LTPDSSPKGLJAPAPQPEEPEDDAESSPDLOAFWFNISAQMPCNAAAPTPLPAVSEDP 1320
DB 1256 LTPDSSPKGLJAPAPQPEEPEDDAESSPDLOAFWFNISAQMPCNAAAPTPLPAVSEDP 1315
QY 1321 TSPPOQLASSKPMNRPSAANPCSPVQFSSTPLAGLAPKRRAGDPGEMPOSPTGLQPKRR 1380
DB 1316 TSPPOQLASSKPMNRPSAANPCSPVQFSSTPLAGLAPKRRAGDPGEMPOSPTGLQPKRR 1375
QY 1381 GRPSKFFKQEQRLVLTQTAQVPPENCSCGWWIPDEMLDAMLKALHPGIREKALHK 1440
DB 1376 GRPSKFFKQEQRLVLTQTAQVPPENCSCGWWIPDEMLDAMLKALHPGIREKALHK 1435
QY 1441 HLNKHRDFLOEYCLRPSADTPEPQLPAFOEGIMSWSPKETYETDLAVLQWVEELEQR 1500
DB 1436 HLNKHRDFLOEYCLRPSADTPEPQLPAFOEGIMSWSPKETYETDLAVLQWVEELEQR 1495
QY 1501 VIMSDQIRGWTCPSDSTREDLAYCEHLSQSDEDITWRGPGREGAPQRKTNPDLAV 1560
DB 1496 VIMSDQIRGWTCPSDSTREDLAYCEHLSQSDEDITWRGPGREGAPQRKTNPDLAV 1555
QY 1561 MRLAALQNVKRYRLREPLWPTHEVVELEKALLSTENGAPEGTTTIEISYEITPRIRWROT 1620
DB 1556 MRLAALQNVKRYRLREPLWPTHEVVELEKALLSTENGAPEGTTTIEISYEITPRIRWROT 1615
QY 1621 LQRCRSHAVCLCLGHLESTAWKSNVKTCLVCKGDNDEFLLCLDGCGRGCHYICHR 1680
DB 1616 LQRCRSHAVCLCLGHLESTAWKSNVKTCLVCKGDNDEFLLCLDGCGRGCHYICHR 1675
QY 1681 PKMEAVPEGDFWCTVCLAQQVEGFTQKPGPKRGQRKSGSYLNFSEGDGRRRRVLLKG 1740

DB 1676 PKMEAVPEGDFWCTVCLAQQVEGFTQKPGPKRGQRKSGSYLNFSEGDGRRRRVLLRG 1735
QY 1741 RESPAAGPRYSEERLSPSKRRRLSRNHHSLDTFCEIILMEMESHDAAWPFKEPVNPLV 1800
DB 1736 RESPAAGPRYSEERLSPSKRRRLSRNHHSLDTFCEIILMEMESHDAAWPFKEPVNPLV 1795
QY 1801 SGYRRIIKNPMDFSTMRELLRGYTSSEEPADALLVFDNCQTFNEDDSEVGKAGHMR 1860
DB 1796 SGYRRIIKNPMDFSTMRELLRGYTSSEEPADALLVFDNCQTFNEDDSEVGKAGHMR 1855
QY 1861 RFPESRWEFYGQKQANL 1878
DB 1856 RFPESRWEFYGQKQANL 1873
RESULT 8
US-10-087-192-663
; Sequence 663, Application US/10087192
; Publication No. US20020182586A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; APPLICANT: Engelhard, Eric K.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
; FILE REFERENCE: 529452000122
; CURRENT FILING DATE: 2002-03-01
; PRIOR FILING DATE: 2000-12-22
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 2059
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 663
; LENGTH: 1586
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-087-192-663
Query Match 69.0%; Score 6777.5; DB 13; Length 1586;
Best Local Similarity 69.3%; Pred. No. 0;
Matches 1323; Conservative 81; Mismatches 128; Indels 377; Gaps 9;
QY 1 MEMEANBANDHFNFTGLPPAPAAAGLAPSPSSGEGLYTNGSPMNFPOQKSLNGDVNVNG 60
DB 24 MEM---EANDHFNFTGLPPAPAAAGLAPSPSSGEGLYTNGSPMNFPOQKSLNGDVNVNG 80
QY 61 LSTVSHTTTSGILNSAPHSSSTSHLHPSVAYDCLMNYQYPSANPGSNLKDPPLLSQFS 120
DB 81 LSTVSHTTTSGILNSAPHSSSTSHLHPSVAYDCLMNYQYPSANPGSNLKDPPLLSQFP 140
QY 121 GGQYPLNGILGGSQPSPSHNTNLRAGSQKFWANGTHSPMGLNFDQSELVDSPPDQNF 180
DB 141 GGQYPLNGILGGSQPSPSHNTNLRAGSQKFWANGTHSPMGLNFDQSELVDSPPDQNF 200
QY 181 -----EVCIGHIDEAAEKEMTSVVAENGTLVCS 210
DB 201 VMPNGPSPFTTSPOTSPMLGSSIQTFAPSDQVSDIHPDEAAEKELTSVVAENGTLVCS 260
QY 211 LELEEXQPELKMCGYNGSVPSVESHQEVSVLVPDPTVSCLDPPSHLPDQLEDTPILSED 270
DB 261 LELEEXQPELKMCGYNGSVPSVESHQEVSVLVPDPTVSCLDPPSHLPDQLEDTPILSED 320
QY 271 SLEPFNSL-APEPVSGGLYGIDDTDELGAEDKPLXDSPVISALDCPSLNATAPSLAD 329
DB 321 SLEPFNSL-APEPVSGGLYGIDDTDELGAEDKPLXDSPVISALDCPSLNATAPSLAD 380
QY 330 DSQSTSIFFASPTSPVVLGEBVLQDNLNNGSDAEQEEMETQSSDPPPSLTQAPAPQS 389
DB 381 DSQSTSIFFASPTSPVVLGEBVLQDNLNNGSDAEQEEMETQSSDPPPSLTQAPAPQS 404
QY 390 STIQLHPATSPAVSFTTSPAVSLVSPVSPASPEISPEVCPAASTVVPVSPVSPASSAV 449

QY 187 -----HPDEAAEKEMTSSVAENGTCGLVCSLELEEXQPELKMCGYNGSVSPVESLHQ 237
DB 265 PRATSSSPAHPKQTLNHNPNFLNAL-----LGNHP-----NG-----VIQSVIQ 307
QY 238 EVSVLPDPTVSCLODPHLPQLEDTPILSEDSLEPENSILAPEVPVSGGLYCIDDTLM- 296
DB 308 EAPL-----ALTTKMQQSKINEN--IAAASSTPSSPNLSTSGRRTPGNTPTVMP 357
QY 297 -----GARDKLPXLDSPVISALDCPSLMNATAFSLADDSTSTSIASPTSPVPL 347
DB 358 SASPIHSGKEKAVANNVNPVK-----QHSHHPAKSLVFEQFGCTDSIPSSKDSN 412
QY 348 GESVLQDINSFDLNGSDABQEMETQSSDFPPLSTOPA-----PDQSTIOLHPAT 398
DB 413 EDEEDDEDEDEDEDDSDQSQSDNSSESDTEGSEEDDDDDKQDQDSDSDTEGEKT 472
QY 399 SPASVPTTSPAVSLVSPAASPEISPEVCPAASVTVSPAVFSVSPASSAVLPAVLEVP 458
DB 473 SWKLNKTT-----SVKSPSMS----- 489
QY 459 LTASVTPSPKASPVTPSPAAAFPTASPAKNDVSVSFLTTADVEEITGEGULTAS--GSDGVNRR 517
DB 490 LTGHSTPNLHIAKAPGA--PAALCESQSOPAFLTSSST-----LTSPPHSGTSKR 541
QY 518 RIATPEEVRPLQHQWRREVRIKKGSHRWQGETWYGPCGKMKQFPVIVIKYLSRNLVHS 577
DB 542 RVTDERELRIPLIEYGWQRETRIRNFGRLQGEVAYYAPCGKRLQYQYVIVIKYLSRNGIMD 601
QY 578 VRREHFSRPMPCVDFFERTDPGLQWVOLSABEISRIQAITGKRGPRNTEKATK 637
DB 602 ISRDNFSKARIVGDFYEARDGQEMQWCLKEEDVIPRIAMEGRGRPNPNDRQAR 661
QY 638 EYPVKRGGRPPKVKITELANKTNDNRLKLEAQE-----TLNEEDKAKTAK 685
DB 662 EESRMRKGRPNVNGNAEFLDNADAKLLRLKLOAEIARQAQIKLLRLKLOEQEQRVAK 721
QY 686 SKKQWR-----QKVORGECLTTIOGARNRKQBTSLKH-----KEAKKXSAEK 733
DB 722 EAKQQAIAAABEKKEQEKIMQKEKIKRIQIRMEKELRAQOILEAKKKKEEAAN 781
QY 734 GK--TKQEKLEK-----VKEEKEKVKM-----KEEVEVTKAKPACKAD 771
DB 782 AKLLEAKRIKEKEMRQOAVLLGHQERRRRQHMLMKAMEARKKAEKERLKOEKDE 841
QY 772 KTLATQRLERKQOMILEMKKPTEDMCLTDHOPDPFSRVPGLTLPSPAFSCLTIV 831
DB 842 KRLNKRLEQRLEWAKELKKNEDMCLADQKLPFLPRIGLVLSGSTFSDCLMVV 901
QY 832 EFLHFGKVLGPDPAKDVPSLGVLEQLLCOGDSLGEVQDILLVRLKKAALHDPGPPSYCQ 891
DB 902 QFLRNFVKLVGDVNDVFNLSVLQEGLLNIGDSMGEVQDILLVRLLSAAVCDPGLITGYK 961
QY 892 SILIIGEKVSEIPLTRDNVSEILRFLMAYGVXPALCDRLRTPQOAPPOQAANVLAPP 951
DB 962 AKTALGEHLNNGVNRDNVSEILOTFMEAHCGQTELTSLTKFAQHTPAQKASVLAFL 1021
QY 952 VHELNGSLTLINEIDKLTESMSSYKKNWIVGRLRLKTVLAKTRSE-----VEMGRP 1007
DB 1022 INELACSVSEIDKNIDYNSNLARDKRWVEGKRLKURIHAKTKGRDTSGGIDLGE 1081
QY 1008 EECGL-----RRSSRIME-----ETSGMEEBEEESIAAVPGRGR--RDGEV 1049
DB 1082 QHPLTPTPGRRRRKGGSDYDDDDDDSDQDDEDEDEDKEDQKKTIDCEDEDE 1141
QY 1050 DATASSIELEROIKLSKROLFFRKLLHSSQMLRAVSLGODRVRRYWJLPYLAGIFV 1109
DB 1142 GDQAASVBELEKQIEKLSQSQYRKLFDASHLSRSMFGPDRYRYYWILPRCGGIFV 1201
QY 1110 EGTEGNLPEEVIKETSLSKVAHASLNPAFLFSMKMELAGSNTTASSPARARS--PLKT 1168
DB 1202 EGMESGELEB--IAKERKLLKAEVSQIKKEEMF-----ETSGDSLNCNSNTDCEQEDUKE 1256
QY 1169 KPG-----FMQ--PRHFKSPVRGQDSEQPAQLOPEAQLHVPAPQPOLOQLQLOSHKGFLEQE 1224

DB 1257 KDNTNLFLQKPGSFKLSKLE-----VAKMPPESEVMTF-----KPNAGAN 1298
QY 1225 GSPLSLGOS--QHDLS--QSAFLSWLSQTOSSHSSLLSSVL--TPDSSSPGKL--DPAPSOPP 1278
DB 1299 CCTLISYQNSGKHSLSGSVS-----TATQSNVEKADSNLFWTSSGFGKFSPLPND-- 1350
QY 1279 EEPDEAESPDLOQAFWNISQAQPCN-----AAPTPLAVSEDDQPTSPQO 1326
DB 1351 -----QLLKLTLEKNRQWFSLLPRTPCDDTSLTHADMSTASILVTQSQPPSPSPPTAP 1405
QY 1327 LASSKPMNRPSAANP--CSPVQF--SSTPLAGLAKRRAGDPG-----ENPQSP 1372
DB 1406 LGSS--AQNPVGLNPFALSPLOQVGGVSMGL--QFCGWPTGVVTSNIPFTLSVPSLGS 1460
QY 1373 GLGQPKRRGR-----PPSKF--FKOMEQRYLTQTA-----QpVPEPCS 1410
DB 1461 GLGSENGNSFLTSNVASSSESPPVQNEKATSAQPAAVEKAPVDPPSPKPIEEQOF 1520
QY 1411 GWWIPDPEMLDAMLKALHPRGIREKALHKLHNRDLOEVLCLPSADPIPEPQOLPAF 1470
DB 1521 GWRIIDPEDLKALLKVLHLRGIREKALQKQIKQHLDYITQACLKNKDVAIELNENEEN 1580
QY 1471 Q---EGMSWSPKETVETDLAVLOWVEELQORVIMSDLOIRGWTCTSPDSTREDLAVCE 1527
DB 1581 QVTRDIVENWVEEQAMEMDLSVLQVEDLERRVASASLOVKGMWCPPEPASEREDLVYPE 1640
QY 1528 HLSDS---QSDITWRGPGREG--APQKTTNPDLDAVMRLAALQONVKKRYLREPLWPT 1582
DB 1641 HKSTFKLCKEHDGFTGEDESSAHALERKSNPLDIAVRLADLERNIERR----- 1691
QY 1583 HEVLEKALLSTPGAPEGTTTISEYITPRIRIWRQTLQRCRSHAHVCLCLGHLSERSIA 1642
DB 1692 -----IEEDLAPGLRVRRALSEARSAAQVALCICQLOQLKSIA 1728
QY 1643 WEKSVNKVTCVLCRGDNDEFLLLCDGDCRCHYCHRPKMEAVPEGDWFCVCLAQ--- 1699
DB 1729 WEKSIMKYVICIRKGDNEELLCDGCKGCHYCHRPKITTPDGMWFCPACTAKASG 1788
QY 1700 -----QVBEGETQKPGFPKRGOK-----RKSGYSLNFSSEGDRRRRLVK 1739
DB 1789 QTLKIKLHVAGKKTNE---SKGKKVTLTGDTDEDEDSASTSSSLKRGKNDLQKRM--- 1842
QY 1740 GRESPAAGPRYSEBRLSPSKERRLSMRNHHSDLTFCFCEILMEMESHDAAPFEPVNPRL 1799
DB 1843 -EENTSINLSQESFTSVKPK----RDSKDLALCSMILTEMETHEDAMFLLPVNLKJ 1897
QY 1800 VSGYRRIIKNPMDFSTWRERLLRGYTSSEFAADALLVFONCQTFNEDDDSEVGKAGHIM 1859
DB 1898 VPGYKVIKPMDFSTIREKLSGQYPNLETALDVLVFNDCETFNEDDDSDIGRAGHNM 1957
QY 1860 RRFESRWEFEYQ 1872
DB 1958 RYFEKKWTDTFK 1970
RESULT 10
US-10-376-537-21
; Sequence 21, Application US/10376537
; Publication No. US2003022440SAI
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042001
; CURRENT APPLICATION NUMBER: US/10/376,537
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US/09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570


```
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 21
; LENGTH: 1972
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-376-537-21

Query Match      25.3%; Score 2482.5; DB 15; Length 1972;
Best Local Similarity 32.2%; Pred. No. 4.1e-126;
Matches 680; Conservative 329; Mismatches 678; Indels 426; Gaps 68;

Qy      1 MEMEANEAANDHFNFTGLPAPAAAGLKPSPS-----SCEGLTYNGSPNFPQQ 48
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy      43 MESSNSUSD-----SGTSDTSISSEGISSDSDDLDEDEEEDDQSTEESEDSDUSEBAQ 98
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy      49 GKSLLNGDVNVLSTVSHTTTSGILNSAPHSSSTSHLHP--SVAYDCLWNTSYQPSANPG 107
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy      99 HKS-NNQVLLHG---ISDPKADG--QKATERAQEKRIHQPLPLAFE-----SQTHSFQ-- 145
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy      108 SNLKDPPLLSQ-----FGGQVPLNGI-----LGSRRQPS----- 138
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy      146 SQQKQPVLSQQLPPIFQSSQAKESVNNKHTSVIQTGLVSNVKNPPLSLVNOAKKETVMKL 205
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy      139 --PSHNTNLRAGSQKFWANGTHSPMGLNFPDQELYDSPDO-----NFEHVCSGI-- 186
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy      206 IVPSPDV--LKAGNKNWTSBESSLLTSELKSKREQYKQAPLQKQOESSKSUKKVAALSN 264
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy      187 -----HPDEAAEKEMTSVBAENGTLVCSLELEEXQPELKMCGYNGSVPSVSLHQ 237
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy      265 PKATSSSPAHPKQTLNNHNPFLTNAL-----LGNHQP-----NG---VIQSVIQ 307
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy      238 EVSVLVDPPTVSCLDPDHLPDQLEDTPILSEDSLEPPNSLAPBPVSGGLGVIDDTELM-- 296
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy      308 EAPL-----ALTTKTKMQSKINEN--IAAASSTPSPVNLSTSGRRTPGNQTPVP 357
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy      297 -----GAEDKLPXDSPVISALDCPSLNNATAFSLADDSTSTSIASPTSPVPL 347
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy      358 SASPILHQGKEKAVNNVNPVKT-----QHHSHPAKSLVQFRGTDSDIPSSKDSN 412
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy      348 GESVLQDINSFDLNNGSDAEQEMETQSSDFPPSPTQPA-----PDQSSTTQLHPAT 398
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy      413 EDEEEDDEEEDDEEEDDSQSESDSSESOTEGSEEDDDDKQODESDSDEGEKT 472
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy      399 SPAPVPTTSPAVSLVSPAASPEISPEVCPAASTVSPAVFSVSPASSAVLPAVSLEVP 458
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy      473 SMKLNKTT-----SVKSPSMS----- 489
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy      459 LTASVTSPKASPVTSAPAAFTASPANKDVSSFLETTADVETITGEGLTAS--GSGDVMR 517
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy      490 LTGHSTPRNLHIAPGSA--PAALCSQSQAFLGTSSST-----LTSSPHSGTSKRR 541
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy      518 RIATPEEVRLPLQHGWRREVRIKKGSHRWQGETWYGPCKRMKQFPEVIKYLGRNLVHS 577
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy      542 RVTDERELRIPLEYQWQRETRINRFGGLQGEVAYAPCGKKLRQYPEVIKYLGRNGIMD 601
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy      578 VRREHFSRPMVGDFFPEERDTPBGLQWVLSAEIIPSRIOATGKRGPRNTEKATK 637
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy      602 ISRONFSFSAKIRVGDYFYEARDGPOEMQWCLLKEEDVIPRIAMEGRGRPNPDRQAR 661
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy      638 EVPKVRGRGPPKVKITELANKTNDNRPLKLEAQE-----TLNEEDKAKTAK 685
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy      662 BESRRRRKRRPPNVGNNAEFLDNADAKLRLQQAQEIARQAQAKLRLKQKQBOARVAK 721
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy      686 SKKKMR-----QKVRQEGCLTTIOQARNRKRQKQETKSLKH-----KEAKKSKAEXEK 733
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy      722 EAKKQQAIAAAEKEKQEQKIMKQEQEKIKRIQOIRMEKELRAQOILAEKKKKEEAAAN 781
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy      734 GK--TKQSKLEK-----VREKKKVKVM-----KEKEEVTKAKPACKAD 771
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy      782 AKLLEAEKRIKEKMRROQAVLLKHQERRRRQHMMLMKAMEARKKABEKKERLKQEKRDE 841
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

772 KTLATQRRLEERQKQMIILEMKKPTEDMCLTDHQPLPDFSRVPLGLTLPSPGAFSDCLTIV 831
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
842 KRLANKERLEQRLELEWAKELKFPNEDMCLADQKPLPELPRIPLGLVLSGTSFSDCLAVV 901
Qy      832 EFLHSFGKVLGFDPAKOVPSLGVLOEGLCOGDSLGEVQDILLVRLLLKAALHDPDPSPVCO 891
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
902 QFLNFGKVLGFDVNIIDVFNLSVLQEGLLNIGDSMGEVQDILLVRLLSAAVCDPGLITGYK 961
Qy      892 SLKLTGKVKSEIPLTRDNVSEILRCFLMAYGVXPCALCDRLTQPPQAPPOOKAAVLAPF 951
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
962 AKTAGLHLLNVGNVNRNVSEILQIFMEAHCGQTELTESLTKAFQAHTPAKASVLAFL 1021
Qy      952 VHELANGSTLIIINEIDKLTLESMSYKRNKWIVEGRRLRLKTVLAKRTGRSE----VEMGRP 1007
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1022 INELACSKSVSEIDKNIDYMSNLRDKWVVEGKRLKRIIHAKTGKEDTSGGIDLGE 1081
Qy      1008 EECIG-----RRRSRIME-----ETSGMBEEEEESIAAVPGRGR--RDGEV 1049
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1082 QHPLGTPTPGRKRRKGGDSYDDDDDDSDQDQDEDEDEDEKQKKTDCEDDEDE 1141
Qy      1050 DATASSIPELEROIEKLSKQLFRKLLHSSQMLRAVSLGQDRYRRYVWLPYLAGIFV 1109
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1142 GDQAASVEELEKQTEKLSKQSQYRRKLFDAHSLRSVMFGPDRYRRYWILPCGGGIFV 1201
Qy      1110 EGTGNLVPPEVVKETDSLKVAHAASLNLPALFPMKMKELAGSNTTASSPARARS-PLKT 1168
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1202 EGMESGLEB-IAKERELKKAESVQKEMF-----ETSGDSLNCSTNDHCEQKEDLKE 1256
Qy      1169 KPG---FMQ--PRHPKSPVRGQDSEQQAQLQPEAQLHVPAQPQLOLQLOSHKGFLOE 1224
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1257 KDNTNLFQKPGSPSKSLKLE---VAKMPPESEVMTF-----KPNAGAN 1298
Qy      1225 GSPSLGQS--QHDLS--QSAFLSWLSQTSQSHSLSSVL--TPDSSPGKL-DRAPSO 1278
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1299 GCTLSYQNSGKHSLSGVSQ-----TATQSNVEKADSNLNFNTGSSGKGYSLPND-- 1350
Qy      1279 BEPPEDEAESPDLOAFWFNISAQMPCN-----AAPTPLAVSDDQPTSPSQ 1326
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1351 -----QLLTKITEKRWQFSLPRTPCDDTSLTHADMSTASLVTPQSPPSKSPPTAP 1405
Qy      1327 LASSKPMNRPSAAMP--CSPVQF--SSTPLAGLAPKRRAGDPG-----EMPQSP 1372
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1406 LGSS--AQNPVGLNFPFALSPLOVKGGVSMGL---QFCGWPTGVVTSNIPFTLSVPSLGS 1460
Qy      1373 GLQPKRGR-----PPSKF--FKOMEQRYLTOLTA-----QVPYPPMCS 1410
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1461 GLGSEGNNGSFLTSNVASSKSESFPVQNEKATSAQPAAVEVAKPVDFPSPKPIPEEQF 1520
Qy      1411 GWWIIPDPEMLDAMLKALHPRGIREKALHKLHKLNRDPLQEVCLRPSADPIFEPRQLPAF 1470
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1521 GWWIIPDPEMLDAMLKALHKLHKLNRDPLQEVCLRPSADPIFEPRQLPAF 1580
Qy      1471 Q---EGIMSPKPKTYETDLAVLQWVEELQVRVIMSDLQIRGWTCTSPDSTRDLAYCE 1527
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1581 QVTRDIVENVSVEQAMEMDLVLQVQVEDLERRVASASLQVKGWMCPEPASEREDLVYFE 1640
Qy      1528 HLSDS-----QEDITWRGREGEL--APORKYTNPLDLAMRLAALAEONVKRYLREPLWPT 1582
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1641 HKSPFKLCKEHDGFTGEDSSAHALERKSNPLDIAVTRLADLERNIERR----- 1691
Qy      1583 HEVVLKALLSTPGAPEGTTEISYEITPRIRIWRQTLQRCRGAHVCCLGHLERSIA 1642
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1692 -----TEEDIAFGLRVWRRALSEARSAQVALCIIQLOKQIA 1728
Qy      1643 WEKSVNKTCLVCRKGNDEFLLCDGCDRGCHYIYCHRPKMEAVPEGDFWCTVCLAQ--- 1699
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1729 WEKSIMKYVCOICRKGNEELLCDGCDRGCHYIYCHRPKMEAVPEGDFWCTVCLAQ--- 1788
Qy      1700 -----QVEGETQKPGFPKRGQK-----RKSGYSINFESEGGRRRRLVK 1739
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1789 QTLKIKLHVKGKKTNE---SKGKKVTLTGDTDEDEDSASTSSSLKGNKQDLQKKRM--- 1842
Qy      1740 GRESPAAGPRYSERLSPSKRRRLSMRHHSDLTFCETIILMEMESHDAAPFEPVNPRL 1799
```



```

Db      1843  -EINTSINLSKQESTSVKPK-----RDSKDALCSMLTETHEADWFLLPVNUKL 1897
Qy      1800  VSGYRIIKNPDMFTMRRLRGYTSSEEFADALLVFNQCOTFNEDDSEVGKAGHIM 1859
Db      1898  VPGYKKVIKPMDFSTIREKLSSGQYPNLETFLDVLVFNDCNTEFNEDDSDIGRAGNM 1957
Qy      1860  RPFESRWEFFYQ 1872
Db      1958  RYFEKKWTDTFK 1970

RESULT 11
US-10-702-148-21
; Sequence 21, Application US/10702148
; Publication No. US20040063145A1
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042002
; CURRENT APPLICATION NUMBER: US/10/702,148
; CURRENT FILING DATE: 2003-11-05
; PRIOR APPLICATION NUMBER: PRIOR APPLICATION NUMBER: US/09/839,479
; PRIOR FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 21
; LENGTH: 1972
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-702-148-21

Query Match      25.3%; Score 2482.5; DB 15; Length 1972;
Best Local Similarity 32.2%; Pred. No. 4.1e-126;
Matches 680; Conservative 329; Mismatches 678; Indels 426; Gaps 68;

Qy      1 MEMEANEANDHFNFTGLPPAPAAAGLKPPSPS-----SCEGLYTNKSGPMNFPQQ 48
Db      43 MESSNSDSD---SGTSSDTSSEGISSSDSDLEDEDEEDQSTEESEDDSDSESEAQ 98
Qy      49 GKSINGDVNAGLSTVSTHTTSGILNSAPHSSSTSHLHP-SVAYDCLWNTSYQPSANPG 107
Db      99 HKS-NNQVLLHG---ISDPKADG--QKATEKAQEKRIHQPLFLAPE-----SQTHSFQ-- 145
Qy      108 SNLKDPPILSO-----FSGGOYPLNGI-----LGGSRQPS----- 138
Db      146 SQQKQPQLSQQLPPIFOSSQAKESVNKHTSVTQSTGLVSNVPLSLVNOAKKETYNKL 205
Qy      139 --PSHNTNLRAGSQKFWANGTHSPMGLNFDQELYDSPDQ-----NFEVCSGI-- 186
Db      206 IVPSPDV-LKAGNKVTSSESSILLTSELRSKEQYKQAFPSQLKKQESSKSLKVI AALSN 264
Qy      197 -----HDEAAAEKMTSVVAENGTLVCSLEEXQPELKMCGYNGSVPSVSLHQ 237
Db      265 PRATSSSPAHPKQTLNHNHPFLTNAL-----LGNHQP-----NG---VIQSVIQ 307
Qy      238 EVSVLVPPTVSCLDPPSHLPDQLEDTPILSEDSLEPNSLAPBPVSGLYGIDDITELM- 296
Db      308 EAPL-----ALTTKTKMOSKINEN--IAAASSTPSSPVLNLSSTSGRRTPGNQTPVM 357
Qy      297 -----GAEDKLPLDPSVIALDCPSLNNATAFSLADDSDTSITSIFASPTSPVL 347
Db      358 SASPILHSQGEKAVNNVNPVKT-----QHHSHPAKSLVQFRCGTDSDIPSSKDSN 412
Qy      348 GESVLQDINSFDLNGNSDAEQEMETQSSDFPPSLTQPA-----PDQSSTTQLHPAT 398

```

```

Db      413 EDEBEDDEDEDEDEDDSDSESNSSESOTEGSEEDDDDDKOQDESDDTEGSKT 472
Qy      399 SPAPVPTTSPAVSLVSPAPSPISPEVCPAASTVSPAPVSPVSPASSAVLPAVSLV 458
Db      473 SMKLNKTT-----SVKSPSMS----- 489
Qy      459 LTASVTPKAPVTPSPAAAFPTAPANKDVSSPLETTADVEEITGEGULTAS-GSGDVNR 517
Db      490 LTGHSTPRNLHIAKAPGSA-PAALCSESQSFAFTGTSST-----LTSSPHSGTSKR 541
Qy      518 RIATPEEVRLPLOHGWRRREVRIKKGSHRWQGETWYGPCGKRMKQFPEVIKVLSENVLHS 577
Db      542 RVTDERELRIPILEYGQWQRETRIRNPGRLQGEVAYAPCGKKLKQYPEVIKVLSENGIMD 601
Qy      578 VRREHFSFSPMPYGDFFFEEDTPEGLQWVLSAEIIPSRIOAITGKRGPRNTEKAKT 637
Db      602 ISRDNFSFAKIRVGDVEARDGPEQMCLKEEDVIPRIRAMEGREGRPPNPDQRAR 661
Qy      638 EVPKVGGRGPPKVKITELLNKTNDNRPLKLEAQE-----TLNEEDKAKTAK 685
Db      662 EESRMRKGRPPNVGNAEFLDNADAKLLRKLQAEIARQAAQIKLLRKLQKQEQARVAK 721
Qy      686 SKKMR-----QKVQGECLTTTQGAARKKQETKSLKH-----KEAKKSKAEKEK 733
Db      722 EAKQQAImAAEERKQKEQIKIMKQEKIKRIQOIRMEKELRAQQILEAKKKKEBAAN 781
Qy      734 GK--TKOEKLKEK-----VKREKKEKVKM-----KEKEEVTKAKPACKAD 771
Db      782 AKLLEAKRIKEKEMRRQQAALLKHQERERRRQHMMLKAMEARKKAEKERLQEKDE 841
Qy      772 KTLATQRRLEERQKQMILEMKKPTEDMCIJDHQLPDFSRVPLGLTLPSCAFSDCLTIV 831
Db      842 KRLNKRKLEQRLEMAKELKPNEDMCLADQKPLPELPRIPGLVLSGSGTSDCLMVV 901
Qy      832 EFLHSFGKVLGFPDAKQVPSLVLQEGLLCOGDSLGEVODLLVLLKALHDHPGPSVCQ 891
Db      902 QFLNFGKVLGFDVNIIDVPLNSVLQEGLLNTGDSMGEVQDILLVLLSAAVCDPGLITGYK 961
Qy      892 SLKTLGKVSIEPIPLTRDNVSEILRCFLMAYGVXPCALCDRLTQPFQAOOPQOKAAVLFP 951
Db      962 AKTALGEHLLNVGNRDNVSEILQIFMEAHCGQOTELTESLTKAFQAHTPAKKASVLAPL 1021
Qy      952 VHEUNGSTLIINEIDKLTLESMSYKRNKWIVEGLRLRLKTVLAKRTGSE-----VEMGRP 1007
Db      1022 INELACSKSVVSEIDKNIDVMSNLRDKWVVEGKRLRIIHAKTGKRTSGGIDLGE 1081
Qy      1008 EECILG-----RRSSRIME-----ETSGMEEEEEEESIAAVPGRGR--RUGEV 1049
Db      1082 QHPLGTPTPGRKRRKGGSDYDDDDDDSDQDEDEDEDEKEDQKGTDCEDDE 1141
Qy      1050 DATASSIPELERQIEKLSKROLPPFKLLHSSQMLRAVSLGQDRYRRYVWLPYLAGIFV 1109
Db      1142 GDQAASVEELEKQIEKLSKQOYRKLFDASHLSRVWFGPDYRRYRWIPLRCGGIFV 1201
Qy      1110 EGTEGNLVPBEVVKETDSLKVAHAASLNALPALSFMKMELAGSNNTTASSPARAR--PLKT 1168
Db      1202 EGMESGEGLEE-IAKERELKKAESVQIEEMF---ETSGDSLNCSTNDHCEQEDLKE 1256
Qy      1169 KPG---FMQ--PRHPKSPVRGQDSQQAQLOPEAQHLVPAQPOLOLQLOSHKGFLEQE 1224
Db      1257 KDNTNLFQKPGSFKSLKLE---VAKMPPESEVMT-----KPNAGAN 1298
Qy      1225 GSPLSLQGS-QHDLG--QSAPLSMLSOTQSHSLLSSSLV--TPDSSPGKL-DPAPSQP 1278
Db      1299 GCTLUSYNSGKSLGVSQS-----TATQSNVEKADSNLNTGSSGPKGKYSPLND-- 1350
Qy      1279 BEPEFDEAESPDILQAFWFIISAQPCN-----AAPTPLAVSDEQTPSPQ 1326
Db      1351 -----QLLTKLTLEKRNQWFSLLPRTPCDDTSLTHADMSTASLVTPSQSPSPSPAP 1405
Qy      1327 LASKPMNRPSAAMP--CSPVQF--SSTPLAGLAPKRRAGDPG-----EMPQSP 1372

```

Db 1406 LGSS--AQNVPCLNPFALSPLOVKGVSMMGL---QFCGWPTGVVTSNIPFTLSVPSLGS 1460
Qy 1373 GLGQPKRGR-----PPSKF---FKQMEQRYLTQLTA-----OPVPPMCS 1410
Db 1461 GLGSEGNNGFLTSNVASSSESPVQNEKATSAQPAAVEVAKVPDPSPKPIPEMQF 1520
Qy 1411 GWWIPDPMDLAKLALHPRGIREKALHKLHNRDPLQBVCLRPSADPIFEPRQLPAF 1470
Db 1521 GWWRIIDPEDLKALLKVLHLRGIREKALOKQIQKHLDYITQACLKNDKVAIIENENEEN 1580
Qy 1471 Q---EGIMSWPKTYETDIAVLQWVEELQORVIMSDQIRGWTCPSPDSTREDLAYCE 1527
Db 1581 QWTRDIVENWSEQAMEMOLSVLQOVEDLERRVASASLQVKGWCPPEASERDLYVEF 1640
Qy 1528 HLSDS---QEDITWRGPRGL-APQKTTNPLDLAVMRLLAALQNVKRYLRLEPLWPT 1582
Db 1641 HKSFTKLCKEHDGFTGEDESSAHLERKSNPLDIAVRLADLERNIERR-----1691
Qy 1583 HEVVLKALLSTPNGAPEGTTTEISYETTPRIRIWRQTLQRCGAHVCCLGHLERSIA 1642
Db 1692 -----TEEDIAPGLRVWRALSEARSAQAQVALCIQIQKSLA 1728
Qy 1643 WEKSNVKTCLVCRKGNDEFLLLCDGCDRCHYCHRPKNMEVPEGDFWCTVCLAQ---1699
Db 1729 WEKSNVKTCLVCRKGNDEFLLLCDGCDRCHYCHRPKNMEVPEGDFWCTVCLAQ---1699
Qy 1700 -----QVEGETQKPGFKRGK-----RKSGYSLNSEGGRRRVLLK 1739
Db 1789 QTLKIKLHVKKKTNE---SKGKKVTLTGDTDEDSASTSSSLKRGKNDLQKRM---1842
Qy 1740 GRESPAAGPRYSERLSPSKRRLLSMRHHSDLTFCETILMEMESHDAAPFEXEPVNPRL 1799
Db 1843 -EENTSINLSQESFTSVKPKF---RDDSKDLALCSMLTETHEDEWFFLLPVNLKL 1897
Qy 1800 VSGYRIIRKPMDFSTWRERLLRGYTSSEFAADALLVFNDCOTFNEDDSEVGKAGHIM 1859
Db 1898 VPGYKVIKPMDFSTIREKLSSGQYPNLETFALDVLVFNDCOTFNEDDSDIGRAGHNM 1957
Qy 1860 RPFESRWEFFYQ 1872
Db 1958 KYFEKKWTDTFK 1970

RESULT 12

US-09-839-479-71

; Sequence 71, Application US/09839479

; Publication No. US2002003979A1

; GENERAL INFORMATION:

; APPLICANT: Jones, Michael H.

; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR

; FILE REFERENCE: 06501-042002

; CURRENT APPLICATION NUMBER: US/09/839,479

; CURRENT FILING DATE: 2001-04-20

; PRIOR FILING DATE: 1999-10-15

; PRIOR FILING DATE: 1999-10-15

; PRIOR FILING DATE: 1998-04-17

; PRIOR FILING DATE: 1997-10-24

; PRIOR FILING DATE: 1997-04-18

; NUMBER OF SEQ ID NOS: 72

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 71

; LENGTH: 1969

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-839-479-71

Query Match

Best Local Similarity 24.8%; Score 2436; DB 9; Length 1969;

Matches 683; Conservative 320; Mismatches 635; Indels 468; Gaps 74;

Qy 51 SLANGVNNGVGLSTVSHTTTSGILNSAPHSSSTSHLHHPVSVAYDCLMWNYSQYSPANPGSNL 110
Db 46 SSNSD---SSGTSSTSTSEGI-----SSSDSDLEDEDEEEOQSIESEDDDDSDSESEA 97
Qy 111 KDPFLLSQFGGQYPLNGIILGGSQSPSPSHNTNLRAGSKFWANGTHSPNGLNFDQSGL 170
Db 98 QHK-----SNNQVLLHGI---SDPKADGQKATEKAQEKRI-----HQPPLAFESQT- 141
Qy 171 YDSPDQNF-EVCSGTHP-----DEAAEKEM---TSVAENGTLVCSLELEXQPELK 221
Db 142 -HSFQSOOKQPVLSLQQLPFIFQSSQAKBESVKNHTSVI--QSTGLVSNVKNPLSLVNOAQ 198
Qy 222 MCGYNGSVPSVESLHQVSVLVPDPTV-----SCLDDPSHLPOLEDPTILSDELSLEPF 275
Db 199 KETY-----MKLIVPSDVLKAGNKTSESSLTSEL-----RSKREQY 238
Qy 276 NSLAPFVSGLYGIDTDLGMAEDKLPLXDSVISALDCPSLNNATAFSLDADSOTST 335
Db 239 KQAPP-----SQLKKQESSKSL--KKVIAALSNNPK-----AT 268
Qy 336 SIFASPTSPVILGESVLQDHSFD--LNN--GSDAEQEEMETQSDPPPSLTQAPDOSS 390
Db 269 S--SSPAHP---KOTLENNHPNPLTNALLGNHPNGVIOQSVIOEAPLATTTKQWQSK 322
Qy 391 TIQLHPATSPAVSTTSPAVSLVSPASPEISPEVCPAASTVV-----SPAIVSVVS-- 443
Db 323 INENIAAASS--TPFSSP-VNLSTSGRTFQNTQFVMPSPASPIILHSQKEKAVNNVNPV 379
Qy 444 -----PASSAV-----LPA-----452
Db 380 KTOHHHPAKSLVFOFRGTDSIDIPSSKXSDSDEDEDEDEDEDEDEDEDEDESDS 439
Qy 453 -----VSLEVLPTASVTSFKAS-----469
Db 440 SNSESDTEGSEEDDDKQDQSDSDTEGETSKMLKNTTSSKSPSLTGHSTPRNLHI 499
Qy 470 ---PVTSPAA-AFPTASPAKNDVSSFLETTADVBEITGEGLTAS--GSGDVMRRRIATPEE 524
Db 500 AKAPGSAPALCSSESQSPA-----FLGTSST-----LTSSPHSGTSKERRVTDERE 546
Qy 525 VRLPQHGWRRVRAIKGSHRWQGETWYQPCGKRMKOPPEVITKYLNRNLVHSVRRBHS 584
Db 547 LRLPLEYQWQRETRIRNFGRLQGEVAYYAPCGKKLQYPEVIKYLNRNIMDISRDNFS 606
Qy 585 FSPMPYVDFPEERDTEGLQWVLSAEIIPSRIOAITKGRPRNTEKATKVPKVR 644
Db 607 FSAKIRVDFYEARDGPOEMQWCLLKEEDVIPRAMEGRGRPRPNDRQARESMRR 666
Qy 645 GRGRPPKVKITELINKTDRPLKKLEAQET-----LNEDKAKIAKSKKQWR-- 691
Db 667 RKGRPFNVGNAEFLDNADAKLLRLKLOQEARQAAQIKLLRLKLOQEARVAKEAKQQA 726
Qy 692 ---QKVORGECLTTIQOARNKKTQETSLKH-----XBAKKKSAKEKKGK--TKQE 739
Db 727 MAABEKKQKEQIKIHQKQEKIKRIQOIRMEKELRAQOILEAKKKKEEANAALLAEK 786
Qy 740 KLKEK-----VKREKKEKVM-----KEKEVTKAKPACKADKTLAQR 779
Db 787 RIKEREMVRROQA VLLKKQERRRRQHMLMKAMEARKKAEKERLQKEKDEKLNKERK 846
Qy 780 LEERQKQOIMLEEMKPTEDMCLTDHOPDFSVPGVGLTLPSGAFSDCLTVEFLHSPGK 839
Db 847 LEQRLELEMAKELKKFNEDMCLADQKPLPELPIPGVLSGSTFSDCLMVVQFLRNGK 906
Qy 840 VLGFDPADVPVSLGVLOEGLLCQ--GDSLGEVQDILLVRLKALHDHPGPPSYCSQLKILGE 898
Db 907 VLGFVDNIDVPLSVLQEGILLNIGDSMGVQDILLVRLLSAAVCDPGLITGYKAKTALGE 966
Qy 899 KVSEIPLTRONVSEILRCFLMAYGVXPALCDRLTQPOAQPOPOKAAVLAFPPVHELNGS 958
Db 967 HLLNVGNVRDNVSEILQIFMEAHCGQOTELTESLTKAFQHTPAQK-AVLAFLINELACS 1025
Qy 959 TLIINEIDKTLESMSYRKNKWIVEGRLLRLKTLAKRTGRSE----VEMGRPSECLG-- 1012

Db	1026	KSVSEIDKNIDYMSNRDRKNVVEGKRLKURIITHAKTKGRDTSGGIDIGERQHPJGTP	1085
Qy	1013	-----RRSSRIME-----ETSGWEEBEEESIAAUPGRRGR--RDEGEVDATASSI	1056
Db	1086	TPGRKRRKGGSDSYDDDDDDSDQDQDEDEDEKQKKTDCIDEDECDQAASV	1145
Qy	1057	PELEROQTEKLSKROLFFRKLLHSSQMLRAVSLGQDRYRRYVWLPYLAGIFVEGTGNL	1116
Db	1146	BELEKQTEKLSKQOSQYRRKLFDAHSLRSYMFQDRYRRYVWLPYCGGIFVEGMESGE	1205
Qy	1117	VPEBIVKETTDSLKVAASHALNPALFSMKWELAGSNITASSPARASR-PLKTKPG--F	1172
Db	1206	GLEB-IKEREKLLKAASVQ:KEEMF-----ETSGDSLNCNSNTHCEQEDLKCKDNLNF	1260
Qy	1173	MQ-PRHFSPVVRGDSQPQAOQAOPEAQLHVPAQPQLOLQLOSHKGFLEQSGPLSLG	1231
Db	1261	LQKPGSFKLSKLLLE---VAKMPPESEVMTP-----KPNAGANGCTLSYQ	1302
Qy	1232	QS-OHDLIS--QSAFLSWLSQTOSSHLSLLSSVL--TPDSSPGKL-DPAQOPPEEPEDDE	1285
Db	1303	NSGKHSLSGSVQS-----TATQSNVEKADSNLNFNTGSSPGKFYSPLPND-----QL	1349
Qy	1286	AESSPLQDAFWFNISAQMPCN-----AAPTPLAVSDEQTPSPQOLASSKPM	1333
Db	1350	LKTLUTEKNRWQFSLPRTPCDDTSLTHADNSTASLVTPQSPPKSPSPPTAPLGSS--A	1407
Qy	1334	NRPSAANP--CSPVQF--SSTPLAGLAPKRAGDFG-----EMPOSTGLGQPKR	1379
Db	1408	QNPVGLNPFALSPLOVKGGVSMGL---QFCGMPGVVTSNIPPLSVPSLSGLGJLSEG	1464
Qy	1380	RGR-----PPSKF--FKQMEQRYLTQJTA-----QVPPPEMCSGWMIIPD	1417
Db	1465	NGNSFLTSNVASSSESFPQNEKATSAQAAVEAKVPDVPFPSPKPIPEEMQFGWRIID	1524
Qy	1418	PEMLDAMLKALHPRGIREKALHKLHKNHRDFOEVLCLRPSADPIPEPQRLPAQ---	1474
Db	1525	PEDLKALKVLHURGIREKALQKQIQKHLBYITQACLUKNKDVAIELNEENEQVTRDIV	1584
Qy	1475	MSWSPEKTYETDLAVLQWVEELEQRVIMSDLOIRGWTCPSPDSTREDLAYCEHLSDS--	1532
Db	1585	ENWSVEQAEEMDLSVLQOVEDLERRVASASLQVKGMWCEPEASEREDLYVEHKSFTKL	1644
Qy	1533	--QEDITWRGPGREGL-APORKTNPJLDLAVMLRLAALQNVKORYLREPLWPTHEVVLEK	1589
Db	1645	CKEHGDEFTGEDESSAHALERKSDNPJDLIAVTLADLERNIERR-----1688	
Qy	1590	ALLSTPNGAPEGTTTETSIYEITPRITWRQTLQCRSAAHVCLIGHLEBSIAWEKSVNK	1649
Db	1689	-----FEEDIAPLGRVWRRLSEARSAAQVLCIQLOKSIaweKSIMK	1732
Qy	1650	VTCLVCRKGDNDLFLLLCDGCDRGCHTYCHRPKMEAVPEGDWFCVCLAQ-----	1699
Db	1733	VYCOICRKGNEELLLLCDGCDKGCHTYCHRPKITTPDGMWFCPACIAKASGQTLKIKK	1792
Qy	1700	-QVEGETQKPGPPKRGQK-----RKSGYSLNFPSEGGRRRRVLLKGRSSPAA	1746
Db	1793	LHVKGKKTNE---SKKGKVTLTGDTDEDEDSASTSSSLKRGKNDLOKQKKN-----	1845
Qy	1747	GPYSEERLSPSKRRRLSMENHSHDLTFCEIILWEMESHDAAPFXEPVNPRLVSGVRR	1806
Db	1846	NLSQESFTSVKKPK-----RDDSGLDALCDSMIITEMETHEDAWPFLFLVNLKLVPGYKV	1901
Qy	1807	IKNPMDFSTMRILLRRCGYTSSEBAADALLVFDNCOTFNEDDESVEGKAGHIMRRFPESR	1866
Db	1902	IKKPMDFSTIREKLSSQYFNLETFALDVLVFDNCETTFWEDDSDIGRAGHNWRKYFEK	1961
Qy	1867	WERFYQ	1872
Db	1962	WTDTFK	1967

```

US-10-376-537-72
; Sequence 72, Application US/10376537
; Publication No. US20030224405A1
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION, TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042001
; CURRENT APPLICATION NUMBER: US/10/376,537
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US/09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 72
; LENGTH: 1969
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-376-537-72

Query Match      24.8%; Score 2436; DB 15; Length 1969;
Best Local Similarity 32.4%; Pred. No. 1.4e-123;
Matches 683; Conservative 320; Mismatches 635; Indels 468; Gaps 74;

QY 51 SLNGDVNVNGLSTVHTTTSSTGILNSAPHSSSTSHLHHPSVAYDCLMNYSOVPSANPGSNL 110
Db 46 SSND-----SSGTSSDTSSEGI-----SSSDSDLEDEEBEEOQSIEESDDSDSSEA 97
QY 111 KDPPLLSQFGGGQYPLNGILGSRQRPSPSHNTLNLRAGSKFWANGTHSPMGLNFDSEQEL 170
Db 98 QHK-----SNNQVLLHGI---SDPKADGQKATEKAQEKRI-----HQPLPLAFESQT- 141
QY 171 YDSPDQNF-EVCSGHP-----DEAAEKEM---TSVVAENGTGLVCSLESEEXQPELK 221
Db 142 -HSFQSQKQPVLSQQLPFIPOSSQAKESVNKHTSVI--QSTGLVSNVKPLSLVNOAK 198
QY 222 MCGYNGSVPSVESHQBSVLVDPDVTY-----SCLDDPSHLDPLEDDTPILSEDSLPEF 275
Db 199 KETV-----MKLIVSPDVLKAGNKTSESSILLTSEL-----RSKREQY 238
QY 276 NSLAPEVPSGGLYGIDDTELMGAEADKLPLXDSPVISALDCPSLNNATAFSLADDSTST 335
Db 239 KQAPP-----SQLKQESSKSL--KKVIAALSNPK-----AT 268
QY 336 SIFASPTSPVILGESVLQDNSFD--LNN-----GSDAEQEEMETQSSDPFSLTQAPDOSS 390
Db 269 S--SSPAHP-----KQTLNNHPFLFNALLGNHPQNGVIOSVIQEAPLALTTKTKMQSK 322
QY 391 TIQLHPATSPAVGPTTSPAVSLVSPAPASPEISPEVCAPAASTVV-----SPAVFSVVS-- 443
Db 323 INENIAAASS--TFPFSFP-VNLSTSGRTGNTQTPVMPSPASPIHLSOGKEKAVSNVNPV 379
QY 444 -----PASSAV-----LPA-----452
Db 380 KTOHHSHPAKSLVEQFRGTDSDIPSSKSDSDSNEDEEDDEDEDEDDSDSDSSED 439
QY 453 -----VSLEVP LTASVTSPKAS-----469
Db 440 SNSESDTGSEEBDDDDKOQESDSDTGEKTSBMKLNKTTTSSKSPMSLTGCHSTPRLNHI 499
QY 470 ---PVTSPPAA-APPTASPAANKOVSSFLFETADVEEITGEGLTAS--GSGDVMRRRIATPEE 524
Db 500 AKAPGSAPAALCSQSQA-----FLGTSSST-----LTSPSHSGTSKRRRVTDERE 546
QY 525 VRPLQHGWRREVRIRKKGSHRWGETWYTGFCGRMKMQFPEVTKYLSRNLVHVSRRHFS 584
Db 547 LRPLLEYGORETRIRNFGRLQGEVAYVAPCGKRLQYQPEVTKYLSRNGIMDIRNFS 606

```



```
; PRIOR APPLICATION NUMBER: 60/044,039
; PRIOR FILING DATE: May 30, 1997
; PRIOR APPLICATION NUMBER: 60/048,093
; PRIOR FILING DATE: May 30, 1997
; PRIOR APPLICATION NUMBER: 60/048,190
; PRIOR FILING DATE: May 30, 1997
; PRIOR APPLICATION NUMBER: 60/050,935
; PRIOR FILING DATE: May 30, 1997
; PRIOR APPLICATION NUMBER: 60/048,101
; PRIOR FILING DATE: May 30, 1997
; PRIOR APPLICATION NUMBER: 60/048,356
; PRIOR FILING DATE: May 30, 1997
; PRIOR APPLICATION NUMBER: 60/056,250
; PRIOR FILING DATE: August 29, 1997
; PRIOR APPLICATION NUMBER: 60/056,296
; PRIOR FILING DATE: August 29, 1997
; PRIOR APPLICATION NUMBER: 60/056,293
; PRIOR FILING DATE: August 29, 1997
; NUMBER OF SEQ ID NOS: 229
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 94
; LENGTH: 196
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-062-831-94

Query Match      10.5%; Score 1030; DB 14; Length 196;
Best Local Similarity 98.5%; Pred. No. 2.3e-48;
Matches 193; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1683 MEAVPEGDFCTVCLAAQVGEFTQKPGFKRQKRGKSGYSLNFSFGDGRRRRVLLKGRE 1742
Db 1 MEAVPEGDFCTVCLAAQVGEFTQKPGFKRQKRGKSGYSLNFSFGDGRRRRVLLKGRE 60

Qy 1743 SPAAGPRYSEERLSPSKRRRLSMRNHSDLTFCCEIILMEMESHDAAMPFXEPVNPRLVSG 1802
Db 61 SPAAGPRYSEGLSPSKRRRLSMRNHSDLTFCCEIILMEMESHDAAMPFXEPVNPRLVSG 120

Qy 1803 YRRIIKNPMDFSTWRERLLRGYTSSEFAADALLVFDNCOTFNEDDSEVKGAGHIMRRF 1862
Db 121 YRRIIKNPMDFSTWRERLLRGYTSSEFAADALLVFDNCOTFNEDDSEVKGAGHIMRRF 180

Qy 1863 FESRWEFFYQKQANL 1878
Db 181 FESRWEFFYQKQANL 196

RESULT 16
US-10-062-599-94
; Sequence 94, Application US/10062599
; Publication No. US20030195346A1
; GENERAL INFORMATION:
; APPLICANT: Steven M. Ruben, et al.
; TITLE OF INVENTION: 32 Human Secreted Proteins
; FILE REFERENCE: PZ006P1
; CURRENT APPLICATION NUMBER: US/10/062,599
; PRIOR FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: 09/690,454
; PRIOR FILING DATE: 2000-10-18
; PRIOR APPLICATION NUMBER: 09/189,144
; PRIOR FILING DATE: 1998-11-10
; PRIOR APPLICATION NUMBER: 60/044,039
; PRIOR FILING DATE: May 30, 1997
; PRIOR APPLICATION NUMBER: 60/048,093
; PRIOR FILING DATE: May 30, 1997
; PRIOR APPLICATION NUMBER: 60/048,190
; PRIOR FILING DATE: May 30, 1997
; PRIOR APPLICATION NUMBER: 60/050,935
; PRIOR FILING DATE: May 30, 1997
; PRIOR APPLICATION NUMBER: 60/048,101
; PRIOR FILING DATE: May 30, 1997
; PRIOR APPLICATION NUMBER: 60/048,356
; PRIOR FILING DATE: May 30, 1997
```

```
; PRIOR APPLICATION NUMBER: 60/056,250
; PRIOR FILING DATE: August 29, 1997
; PRIOR APPLICATION NUMBER: 60/056,296
; PRIOR FILING DATE: August 29, 1997
; PRIOR APPLICATION NUMBER: 60/056,293
; PRIOR FILING DATE: August 29, 1997
; NUMBER OF SEQ ID NOS: 229
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 94
; LENGTH: 196
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-062-599-94

Query Match      10.5%; Score 1030; DB 14; Length 196;
Best Local Similarity 98.5%; Pred. No. 2.3e-48;
Matches 193; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1683 MEAVPEGDFCTVCLAAQVGEFTQKPGFKRQKRGKSGYSLNFSFGDGRRRRVLLKGRE 1742
Db 1 MEAVPEGDFCTVCLAAQVGEFTQKPGFKRQKRGKSGYSLNFSFGDGRRRRVLLKGRE 60

Qy 1743 SPAAGPRYSEERLSPSKRRRLSMRNHSDLTFCCEIILMEMESHDAAMPFXEPVNPRLVSG 1802
Db 61 SPAAGPRYSEGLSPSKRRRLSMRNHSDLTFCCEIILMEMESHDAAMPFXEPVNPRLVSG 120

Qy 1803 YRRIIKNPMDFSTWRERLLRGYTSSEFAADALLVFDNCOTFNEDDSEVKGAGHIMRRF 1862
Db 121 YRRIIKNPMDFSTWRERLLRGYTSSEFAADALLVFDNCOTFNEDDSEVKGAGHIMRRF 180

Qy 1863 FESRWEFFYQKQANL 1878
Db 181 FESRWEFFYQKQANL 196

RESULT 17
US-09-839-479-64
; Sequence 64, Application US/09839479
; Publication No. US20020039779A1
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042002
; CURRENT APPLICATION NUMBER: US/09/839,479
; CURRENT FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 64
; LENGTH: 175
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-839-479-64

Query Match      8.9%; Score 875.5; DB 9; Length 175;
Best Local Similarity 97.7%; Pred. No. 5.2e-40;
Matches 172; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

Qy 780 LEERQKQOMILEEMKKPTEDMCLTDHQPDPFSRVPGLTLPSCGAFSDCLTIIVEFLHSFGK 839
Db 1 LEERQKQOMILEEMKKPTEDMCLTDHQPDPFSRVPGLTLPSCGAFSDCLTIIVEFLHSFGK 60

Qy 840 VLGFDPADKVPISGLVQEGLLCQSDSLGGEVQDILVRLKKAALHDPGFPSPQCSLKILGEK 899
Db 61 VLGFDPADKVPISGLVQEGLLCQSDSLGGEVQDILVRLKKAALHDPGFPSPQCSLKILGEK 120
```


Qy	900	VSEIPLTRDNVSEILRCLFMAYGXKPCALCDRLRTPQFOAQPPQKAAVLAFPVHEL	955
Db	121	VSEIPLTRDNVSEILRCLFMAYGXKPCALCDRLRTPQFOAQPP-QKAAVLAFPVHEL	175
RESULT 18			
US-10-376-537-65			
; Sequence 65, Application US/10376537			
; Publication No. US20030224405A1			
; GENERAL INFORMATION:			
; APPLICANT: Jones, Michael H.			
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR			
; FILE REFERENCE: 06501-042001			
; CURRENT APPLICATION NUMBER: US/10/376,537			
; CURRENT FILING DATE: 2003-02-28			
; PRIOR APPLICATION NUMBER: US/09/418,710			
; PRIOR FILING DATE: 1999-10-15			
; PRIOR APPLICATION NUMBER: PCT/JP98/01783			
; PRIOR FILING DATE: 1998-04-17			
; PRIOR APPLICATION NUMBER: JP 9/310027			
; PRIOR FILING DATE: 1997-10-24			
; PRIOR APPLICATION NUMBER: JP 9/116570			
; PRIOR FILING DATE: 1997-04-18			
; NUMBER OF SEQ ID NOS: 73			
; SOFTWARE: FastSeq for Windows Version 4.0			
; SEQ ID NO 65			
; LENGTH: 175			
; TYPE: PRT			
; ORGANISM: Homo sapiens			
US-10-376-537-65			
Query Match 8.9%; Score 875.5; DB 15; Length 175;			
Best Local Similarity 97.7%; Pred. No. 5.2e-40;			
Matches 172; Conservative 0; Mismatches 3; Indels 1; Gaps 1;			
Qy	780	LEERQKQOMILEENKKTEDMCLTDHQPLDPFSRVPGILTDSGAFSDCLTIVEFLHSEK	839
Db	1	LEERQKQOMILEENKKTEDMCLTDHQPLDPFSRVPGILTDSGAFSDCLTIVEFLHSEK	60
Qy	840	VLGFDPKADVPVSLGVLOEGLCCQDSLGEVQDILLVRLKAAALHDPGFPSCQSLKILGEK	899
Db	61	VLGFDPKADVPVSLGVLOEGLCCQDSLGEVQDILLVRLKAAALHDPGFPSCQSLKILGEK	120
Qy	900	VSEIPLTRDNVSEILRCLFMAYGXKPCALCDRLRTPQFOAQPPQKAAVLAFPVHEL	955
Db	121	VSEIPLTRDNVSEILRCLFMAYGXKPCALCDRLRTPQFOAQPP-QKAAVLAFPVHEL	175
RESULT 19			
US-10-702-148-64			
; Sequence 64, Application US/10702148			
; Publication No. US20040063145A1			
; GENERAL INFORMATION:			
; APPLICANT: Jones, Michael H.			
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR			
; FILE REFERENCE: 06501-042002			
; CURRENT APPLICATION NUMBER: US/10/702,148			
; CURRENT FILING DATE: 2003-11-05			
; PRIOR APPLICATION NUMBER: PRIOR APPLICATION NUMBER: US/09/839,479			
; PRIOR FILING DATE: 2001-04-20			
; PRIOR APPLICATION NUMBER: US 09/418,710			
; PRIOR FILING DATE: 1999-10-15			
; PRIOR APPLICATION NUMBER: PCT/JP98/01783			
; PRIOR FILING DATE: 1998-04-17			
; PRIOR APPLICATION NUMBER: JP 9/310027			
; PRIOR FILING DATE: 1997-10-24			
; PRIOR APPLICATION NUMBER: JP 9/116570			
; PRIOR FILING DATE: 1997-04-18			
; NUMBER OF SEQ ID NOS: 72			
; SOFTWARE: FastSeq for Windows Version 4.0			
; SEQ ID NO 64			
; LENGTH: 175			
; TYPE: PRT			

Qy	956	NGSTL-----IINEIDKTLSSMS--YRKNKWI VEGRLRLKTVLAK-----	995
Db	739	CGKLLTLVSTRDFIEDYVDILQAKQEPRELKAEQHRKEREEAARIRKRKEEKLKQEQ 798	
Qy	996	-----RTGRSEVEMGRPE-----ECLGRSSRSMETSMGEBEEESIAA 1037	
Db	799	KMKEKQEKLBEDQNRSTADISIGEEEREDFTSIESKDTQEKELDQDMFTEDDDPGSH 858	
Qy	1038	VPGRGRRDGEVDATASSIPELERO-----TEKLSKRQLFRFKKLLHSSQ 1082	
Db	859	KRGRGRKRG-----QNGFKETRQEQINCVTRELLTADBEALKQEHQKELLEKIQ 912	
Qy	1083	MLRAVS-----LGQDRYRRYVLPVLAGIFVEG-----TEGNLYPEEVIKKTDSLKVA 1133	
Db	913	SAIACNIFPLGRDMYRYWIFSPIGLFIEDYSGLTEDMLLP-----	957
Qy	1134	HASLNPAIFSMKVELAGSNTTASSPARARSPLKTKPGFMPORHPFKSPVRQDSEPOQA 1193	
Db	958	-----PQV-----RPSFQNNVQSQD-----	970
Qy	1194	LOPEAQLHVPAPQPOQLQLOSHKHGFLEQSGSPLSGQSHLSQSASFLSWLSQTSHS 1253	
Db	971	-----PQV-----	979
Qy	1254	SLSSSVLTPDSSPCKLPAPSQPPEPEDEABESSPLQAFWNISAQPCNAAPTPL 1313	
Db	980	PLMSESTSNIDQGR-----DHSVQLP-----	1001
Qy	1314	AVSEDOPTSPQQLASSKPMNRPSAANPCSPVQFSSTPLAGLAPKRAGDPGEMPQSP 1373	
Db	1002	-----KPVHKPNR-----	1009
Qy	1374	LQPKRRGRPPSKPFKQMEQRYLTQLTQAQVPPKPCSGMMWIPDPEMLDAMLKALHPRGI 1433	
Db	1010	-----WCFYSSCEQLDQIEALNSRGH 1031	
Qy	1434	REKALHKHKNHRDFOBVCILRPSADPIF-----EPROLPAPQEGIMS--WSPKEKTYE- 1485	
Db	1032	RESALKETLLOEKSRICQAQLARFSEEHFSDKQPDSPKPTYSRGRSSNAYDPQMCABK 1091	
Qy	1486	-TDLAVLQWBELEORVTM-----SDLIQIRGWTCPSPD-----STRE 1521	
Db	1092	QLELRDLFLDIEDRIYOGTLGAIKVTDRHIMSALESRGYELLSEENKENGIIKTVNE 1151	
Qy	1522	DLAYCEHLSDSQEDITWRGPRGREGLAPQKRTNP-----DLAVMLAALAEQNVKR 1572	
Db	1152	DVEEMEIDEQTKVIVKDRLLGIKTETPTSTVSTNASTPOSVSVVHYLAMALFQIEQGIER 1211	
Qy	1573	RYLREPL-----WPTHEVVLEKALLSTPNGAPBGTTEISYEITPRIMWQTLQRCESA 1627	
Db	1212	RFLKAPLDASDGSRSYKTVLDR-----WRESLSSASL 1244	
Qy	1628	AHVCLCLGHLERSIAWKSNNKVTCLVCRKGDNDDEFLLCDGCRGCHYCHRPKMEAVP 1687	
Db	1245	SOVFLHLSTLDRSVIWSIILNARCKIKRKGDAENMVLCGDCRGHTTYCVRPKLKTVP 1304	
Qy	1688	EGDNFCTVCLAQQ-----VEGEFTQKQFPKRGQRKSGVSLN 1725	
Db	1305	EGDNFCEPCRPQRCRRLSFRQPSLESDVEDSMGGEDDVEDDEEGQSEEEVEEVE 1364	
Qy	1726	FSEGD-----GR-----RRVLLKGR-----ESPAAGPRYSE-----	1752
Db	1365	QDEDDSQEBEVSUPFKRGRPOVRUPVKTRGLKSSFSRSGOQQEPGRYPSRSQOSTPKTT 1424	
Qy	1753	-----ERLSPSKRR-----	1762
Db	1425	VSSKTRSLRKINSAPPTETKSLIASRSTRHSHGFLQADVVELLSPPRRKRGKSNAN 1484	
Qy	1763	-----LSMRNHH-----	1770
Db	1485	TPENSPNPNFRVIATKSEQSRSVNIASKLSLQESKRRCKRQSPSPVTLGRSS 1544	
Qy	1771	-----DLTPEIILMEMESHDAAPFEPNPRLYSGYRRIIKNPMDFSTMRERLLRG 1823	
Db	1545	GRQGVHELSAFEQVLVVELVTHDDSWFLKLVSKIQVDPDYDIKKPIALNIIEKVNKC 1604	
Qy	1824	GYTSSBEFAADALLVFNQCOTFNEDDSEVGKAGHIMRRFF 1863	
Db	1605	EYKLASEFIDDIELMFNSCFEYNPRNTSEAKAGTRLQAFF 1644	
RESULT 21			
US-10-376-537-1			
; Sequence 1, Application US/10376537			
; Publication No. US20030224405A1			
; GENERAL INFORMATION:			
; APPLICANT: Jones, Michael H.			
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR			
; FILE REFERENCE: 06501-042001			
; CURRENT APPLICATION NUMBER: US/10/376.537			
; CURRENT FILING DATE: 2003-02-28			
; PRIOR APPLICATION NUMBER: US/09/418.710			
; PRIOR FILING DATE: 1999-10-15			
; PRIOR APPLICATION NUMBER: PCT/JP98/01783			
; PRIOR FILING DATE: 1998-04-17			
; PRIOR APPLICATION NUMBER: JP 9/310027			
; PRIOR FILING DATE: 1997-10-24			
; PRIOR APPLICATION NUMBER: JP 9/116570			
; PRIOR FILING DATE: 1997-04-18			
; NUMBER OF SEQ ID NOS: 73			
; SOFTWARE: FastSeq for Windows Version 4.0			
; SEQ ID NO 1			
; LENGTH: 1674			
; TYPE: PRT			
; ORGANISM: Homo sapiens			
US-10-376-537-1			
Query Match 6.8%; Score 671.5; DB 15; Length 1674;			
Best Local Similarity 19.8%; Pred. No. 1.2e-27;			
Matches 305; Conservative 197; Mismatches 399; Indels 639; Gaps 44;			
Qy	644	RGRGPPKVKITELLNKTDNRPLKLEBAQETLNEDKAKIAKSKMKRQKVRQCELTU 703	
Db	424	RRRGPPK-RIH--ISQEDN-----VANKQTLASYSKATKRDKLLKQBEKMSL 470	
Qy	704	QGOARNKQOETSLKHKEAKKSAEKEGKTKQEKLEKVKKEKKEKVKMKKEEVTK 763	
Db	471	AFEXAKLREKADAL---EAKKKEKDEK---KREELKKIVEEE---ELKKKEE--- 516	
Qy	764	AKPACKADKTLATQRRLEEROKQOMILEEMKKPTEDMCLTDHQPDPFSRVPGLTLP 823	
Db	517	-KERLKVREKEREKLEREKRYVEYLKQNSKPREDMCEDDLKELPPTPVK-TLPP 574	
Qy	824	PSCDLTIVEFLHSGKVLGFDPAKDV---SLGVQLQGLLCQGSLSGEVQDLLVRLKAA 880	
Db	575	FGDALMWLEFLNAPGEL--FDLQDEFDPGVTVLELEALV-GNDSEGPLCELLFFLTA 631	
Qy	881	LHDPGFPSPYCSQSLKILGEKYSIEIPLTRDN-----VSEILRFLMAYGV-- 923	
Db	632	F-----QAIABEEEEVAKEQLTDADTKGCSLSKSLDSDCTLSEILRLHILASG 682	
Qy	924	-----XPALCDRLRTQPPQAQPPQQAALVAFPPVHEL 955	
Db	683	TSANAKYRQKRGFDATDACMELRLSNFSLVKLSSTSVYDLTPGEKMLIL-----HAL 738	
Qy	956	NGSTL-----IINEIDKTLSSMS--YRKNKWI VEGRLRLKTVLAK----- 995	
Db	739	CGKLLTLVSTRDFIEDYVDILQAKQSFRELKABQHRKEREEAARIRKRKEEKLKQEQ 798	
Qy	996	-----RTGRSEVEMGRPE-----ECLGRSSRSMETSMGEBEEESIAA 1037	
Db	799	KMKEKQEKLBEDQNRSTADISIGEEEREDFTSIESKDTQEKELDQDMFTEDDDPGSH 858	
Qy	1038	VPGRGRRDGEVDATASSIPELERO-----TEKLSKRQLFRFKKLLHSSQ 1082	
Db	859	KRGRGRKRG-----QNGFKETRQEQINCVTRELLTADBEALKQEHQKELLEKIQ 912	
Qy	1083	MLRAVS-----LGQDRYRRYVLPVLAGIFVEG-----TEGNLYPEEVIKKTDSLKVA 1133	
Db	913	SAIACNIFPLGRDMYRYWIFSPIGLFIEDYSGLTEDMLLP-----	957
Qy	1134	HASLNPAIFSMKVELAGSNTTASSPARARSPLKTKPGFMPORHPFKSPVRQDSEPOQA 1193	
Db	958	-----PQV-----RPSFQNNVQSQD-----	970
Qy	1194	LOPEAQLHVPAPQPOQLQLOSHKHGFLEQSGSPLSGQSHLSQSASFLSWLSQTSHS 1253	
Db	971	-----PQV-----	979
Qy	1254	SLSSSVLTPDSSPCKLPAPSQPPEPEPEDEABESSPLQAFWNISAQPCNAAPTPL 1313	
Db	980	PLMSESTSNIDQGR-----DHSVQLP-----	1001
Qy	1314	AVSEDOPTSPQQLASSKPMNRPSAANPCSPVQFSSTPLAGLAPKRAGDPGEMPQSP 1373	
Db	1002	-----KPVHKPNR-----	1009
Qy	1374	LQPKRRGRPPSKPFKQMEQRYLTQLTQAQVPPKPCSGMMWIPDPEMLDAMLKALHPRGI 1433	
Db	1010	-----WCFYSSCEQLDQIEALNSRGH 1031	
Qy	1434	REKALHKHKNHRDFOBVCILRPSADPIF-----EPROLPAPQEGIMS--WSPKEKTYE- 1485	
Db	1032	RESALKETLLOEKSRICQAQLARFSEEHFSDKQPDSPKPTYSRGRSSNAYDPQMCABK 1091	
Qy	1486	-TDLAVLQWBELEORVTM-----SDLIQIRGWTCPSPD-----STRE 1521	
Db	1092	QLELRDLFLDIEDRIYOGTLGAIKVTDRHIMSALESRGYELLSEENKENGIIKTVNE 1151	
Qy	1522	DLAYCEHLSDSQEDITWRGPRGREGLAPQKRTNP-----DLAVMLAALAEQNVKR 1572	
Db	1152	DVEEMEIDEQTKVIVKDRLLGIKTETPTSTVSTNASTPOSVSVVHYLAMALFQIEQGIER 1211	
Qy	1573	RYLREPL-----WPTHEVVLEKALLSTPNGAPBGTTEISYEITPRIMWQTLQRCESA 1627	
Db	1212	RFLKAPLDASDGSRSYKTVLDR-----WRESLSSASL 1244	
Qy	1628	AHVCLCLGHLERSIAWKSNNKVTCLVCRKGDNDDEFLLCDGCRGCHYCHRPKMEAVP 1687	
Db	1245	SOVFLHLSTLDRSVIWSIILNARCKIKRKGDAENMVLCGDCRGHTTYCVRPKLKTVP 1304	
Qy	1688	EGDNFCTVCLAQQ-----VEGEFTQKQFPKRGQRKSGVSLN 1725	
Db	1305	EGDNFCEPCRPQRCRRLSFRQPSLESDVEDSMGGEDDVEDDEEGQSEEEVEEVE 1364	
Qy	1726	FSEGD-----GR-----RRVLLKGR-----ESPAAGPRYSE-----	1752
Db	1365	QDEDDSQEBEVSUPFKRGRPOVRUPVKTRGLKSSFSRSGOQQEPGRYPSRSQOSTPKTT 1424	
Qy	1753	-----ERLSPSKRR-----	1762
Db	1425	VSSKTRSLRKINSAPPTETKSLIASRSTRHSHGFLQADVVELLSPPRRKRGKSNAN 1484	
Qy	1763	-----LSMRNHH-----	1770
Db	1485	TPENSPNPNFRVIATKSEQSRSVNIASKLSLQESKRRCKRQSPSPVTLGRSS 1544	

Db 859 KGRGKRG-----QNGKFETTRQEQINCVTRELLTADAEALQEHORKEKELLEKIQ 912
Qy 1083 MLRAVS-----LQODRYRRYVWLPYLAGIFVEG-----TEGNLVPEEVIKKETDSLKVA 1133
Db 913 SAIACNTNIFPLGRDRMYRWYIFPSI PGLFIEEDYSGLTEDMLLP----- 957
Qy 1134 HASLNPALFSMKMELAGSNTTASSPARARSPLKTKPGFMOPRHPKSPVRGQDSQPQAO 1193
Db 958 -----RPSFQNNVQSD----- 970
Qy 1194 LQPEAQLHVPAPQPOLQLQSHKGFLEQEGSPLSLQSDHLSQSAPLSWLSQTSQSH 1253
Db 971 -----PQV----- 979
Qy 1254 SILSSSVLTPDSSPGKLDPAFSPPEPEPEDEAESPDLOAFWFNISAQPCNAAPTPL 1313
Db 980 PLMSESTNIDGPR-----DHSVQLP----- 1001
Qy 1314 AVSEDPPTSPQOLASSPMRPSANPCSPVQFSSTPLAGLAPKRRAGDPGEWPSPTG 1373
Db 1002 -----KPVHKPNR----- 1009
Qy 1374 LQPKRGRPPSKFKQEQRYLTQLTAPVPPMCSGMMWIPDPEDMLDMLKALHPRGI 1433
Db 1010 -----WCFYSCEQDQJIEALNSRGH 1031
Qy 1434 REKALHKLHNRDLOEVLCLRPSADPIF-----EPROLPAFQEGIMS--WSPKEKTYE- 1485
Db 1032 RESALKEYLQEKSRICQALAFSEKHFHSDKPDQSKPYSGRSNAYDPSCMAEK 1091
Qy 1486 -TDLAVLQWVELEORVIM-----SDLOIRGWTCSPD-----STRE 1521
Db 1092 QLELRDLFDLIEDRIYOGTIGAIKVTDRHIWRSALSGRYELLSEENKENGIIKTVNE 1151
Qy 1522 DLAYCEHLSDQEDITWGPREGGLAPQKRTNP-----LDLAWRLAALAEONVYR 1572
Db 1152 DVEEMEIDQKVIKVDRLGLIKTETPTVSTNASTPOSVSVVHYLAWLQIEQGHIER 1211
Qy 1573 RYLRPL-----WPTHEVLEKALLSTENGAPEGTTTISYEITPRIRWQTQRCRSA 1627
Db 1212 RFLKAPLADSGRSYKTVLDR-----WRESLLSSASL 1244
Qy 1628 AHVCLCLGHLERSIAWESVKNVCLVCRKGNDEFLLLCDGCRGCHYCHRPQMEAVP 1687
Db 1245 SQVFLHLSTLDRSVIWSKSIILNARCKICRKGDAENMYLVCDCGRGHHTYCVRPKLKTVP 1304
Qy 1688 EGDWFCYCLAQ-----VEGEFTQKGFPPKRGQKRSYSLN 1725
Db 1305 EGDWFCPCRPKQRCRLSFRQPSLESDVEDSMGDEDDVDGDEEGQSEEEYEVE 1364
Qy 1726 FSEGD-----GR-----RRRVLLKGR-----ESPAAGPRYSE----- 1752
Db 1365 QDEDSQEEEVSLPKRGPQVRLPVKTRGLSSFSRGGQOEGRVPSRQOSTPKTT 1424
Qy 1753 -----ERLSPSKRR----- 1762
Db 1425 VSSKTRSLRKINSAPPTETKSLRIASRSTRHSHGLOADVVELLSPRKRRGRKRSANN 1484
Qy 1763 -----LSMRNHS----- 1770
Db 1485 TPENSPNPNFRVIATKSSQSRSVNIAKLSLQSESKRRCKRQSPPEPVTLGRSS 1544
Qy 1771 -----DLTFCEIILMEWESHDAWPFKEPVNPLVSGYRRI IKNPMDFTMBRLLRG 1823
Db 1545 GRQGVHLSAFAQJUVLVRHDDSWPFLKLSKIQVPDYDIILKPIALNIIREKVNKC 1604
Qy 1824 GYTSSEFAADALLVFNDCQTFNEDDSVKGAGHIMRRFF 1863
Db 1605 EYKLASEFIDDIELMFSNCFEYNPNRNTSEAKAGTRLOAFF 1644

RESULT 22

US-10-702-148-1

; Sequence 1, Application US/10702148
; Publication No. US20040063145A1
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042002
; CURRENT APPLICATION NUMBER: US/10/702,148
; CURRENT FILING DATE: 2003-11-05
; PRIOR APPLICATION NUMBER: PRIOR APPLICATION NUMBER: US/09/839,479
; PRIOR FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1674
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-702-148-1

Query Match 6.8%; Score 671.5; DB 15; Length 1674;
Best Local Similarity 19.8%; Pred. No. 1.2e-27;
Matches 305; Conservative 197; Mismatches 399; Indels 639; Gaps 44;
Qy 644 RGRGRPKVKITELLNKTNDNRPLKLSAQETLNEEDKAKIAKSKKMKRQKVORGECLTTI 703
Db 424 RRRGRPPK-RIH--ISOEDN-----VANKQTLASYSRKATKERRDKLLKQEMKSL 470
Qy 704 QGQARNRKQKTSKHKHKAKKSAEKECKTKOEKLKEVKREKKEKVKMKKEEYVK 763
Db 471 AFEKAKURKADAL---EAKKKEKDEK---KRELUKKIVEE---RLKKEE--- 516
Qy 764 AKPACKADKTLATORRLEERQKQMILEEMKKPTEDMCLTDHQLPLDFSRVPLGLTPSGA 823
Db 517 -KERLKVREKEREKREKEKRYEYLKQWSPREDMECODLKELPEPTPVK-TLPPPEI 574
Qy 824 FSDCLTIVFHSFGKVLGFPDPAKDV---SLGVLOEGLLCQGSLSGVQDILLVRLKAA 880
Db 575 FGDALMVLEFLNAGEL--FDLQDEFDPGVTLVELEALV-GNDSEGPLCELLFFFLTAI 631
Qy 881 LHDFGFSYCQSLKILGEKYEIPLTRDN-----VSEILRCFLMAYGV-- 923
Db 632 F-----QAIABEEVEVAKEQLTDADTKGCSLKLSDLSDCTLSEILRLHILASGADV 682
Qy 924 -----XPALCDRLRTQPPQAQPPQQAALVAPPVHEL 955
Db 583 TSANAKYRKGGFDATDDACMELRLSNPSLVKLSSTSVYDLTPGEMKIL-----HAL 738
Qy 956 NGSTL-----IINEIDKLTESMS--YRKQKWIIVEGRRLRLTKTVLAK----- 995
Db 739 CGKLLTLVSTRDFIEDYVILRQAKQBFRELKAEQHRKEBEAARIRKREKELKQEQ 798
Qy 996 -----RTGSESEVMGRPE-----ECLGRRSSRIMSETSCWEEEEEESIAA 1037
Db 799 KMKSKQEKLEDEQORNSTADISIGEEEREDFTSIESKDTQEKELDQDMFTDEDDPGSH 858
Qy 1038 VPGRRGRDGEVDATASSIPELERQ-----IEKLSKQLFRFKLLHLSQ 1082
Db 859 KRGRGRKG-----QNGFKEFTQEQINCVCVTRRELLTADAEALQEHORKEKELLEKIQ 912
Qy 1083 MLRAVS-----LGODRYRRYVWLPYLAGIFVEG-----TEGNLVPEEVIKKETDSLKVA 1133
Db 913 SAIACNTNIFPLGRDRMYRWYIFPSI PGLFIEEDYSGLTEDMLLP----- 957
Qy 1134 HASLNPALFSMKMELAGSNTTASSPARARSPLKTKPGFMOPRHPKSPVRGQDSQPQAO 1193
Db 958 -----RPSFQNNVQSD----- 970

```
QY 1194 LQPEAQLHVPAPQOPOLQLOLQSHKGFLEQEGSPLSLQSQSHLSQSFAFLSWLSQTSQSHS 1253
Db 971 -----PQV-----
QY 1254 SLLSSSVLTDPSSPGKLDPAQSOPPEPEDEAESSPDLOAFWFNISAQMPCNAAPTPL 1313
Db 980 PLMSESTNIDQGR-----DHSVQLP-----
QY 1314 AVSEDOPTSPQOLASSKPMRPNANPCSPVQFSSTPLAGLAPKRRAGDGPMPQSTG 1373
Db 1002 -----KPVHKPNR-----
QY 1374 LQPKRRGRPPSKFKQEQRYLTQLTAQVPPEMCSGMMWIPDPEMLDMLKALHPRGI 1433
Db 1010 -----WCFYSSCEQLDQIEALNSRGH 1031
QY 1434 REKALHKLHNRDFOEVLCLRPSADPIF-----EPRLPAFQEGIMS--WSPKEKTYE- 1485
Db 1032 RESALKETLLQEKSRICQAARFSEKPHFSDKQPDPSKPTYSRGRSSNAYDPQMCAEK 1091
QY 1486 -TDLAVLQWVELEQRVIM-----SDLIQIRGWTCPSPD-----STRE 1521
Db 1092 QLELRDRDPLDIEDRIYQGTGALIKVTRDHIWRSALSGRYELLSEENKENGIIKTVNE 1151
QY 1522 DLAYCEHLSDSQEDITWRGPGREGGLAPQKRTNP-----LDLAVMLAALAEONVKR 1572
Db 1152 DVEWEIDEQTVKIVKDLGIKIKTETPSTVSTNASTPOSVSUVHYLAMALFOIEQGIER 1211
QY 1573 RYLREPL-----WPTHEVVLKALLSTPNAGEGTTTIEISYEITPRIMRQTLQRCRSA 1627
Db 1212 RFLKAPLDASDGRSYKTVLDR-----WRESLSSASL 1244
QY 1628 AHVCLGLHLSRTAWESVNVKVTCLVCRKGDNDDEFLLLCDGCRGCHYIHRPMEAVP 1687
Db 1245 SQVFLHLSTLDRSVIWSKSIILNARCKICRKGDAENMYVLCDCGCRGHHTYCVRPKLKTVP 1304
QY 1688 EGMDFCTVCLAQO-----VEGEFTQKPGFKRGOKRKGYSLN 1725
Db 1305 EGMDFCEPCRKQRCRLRSFRQPSLESDEDEVDMSGGDEVDGDEEGQSEEEVE 1364
QY 1726 FSEG-----GR-----RRVLKGR-----ESPAAGPRYSE----- 1752
Db 1365 QDEDDSQEEVEVSLPKRGKQVRLPVKTRGLSSFSRGGQEQEPGRYPSSQOSTPKTT 1424
QY 1753 -----FELSPKRR----- 1762
Db 1425 VSKTGRSLRKINSAPPTETKSLRIASRSTRSHGPLQADVVELLSPRKRRGRKGSANN 1484
QY 1763 -----LSMRNHS----- 1770
Db 1485 TPENSPNPNFRVIATKSSQSRSVNIASKLSLQSESKRRCRKQSPSPVTLGRSS 1544
QY 1771 -----DLTFCIEILMESHDAWPFEPVNPRLVSGYRRIIKNPMDFSTMRERLLRG 1823
Db 1545 GRQGVHLSAFEQLVVELVDRHDSWPLKLVSKIQVPDYDIIKKPIALNIIREKVNKC 1604
QY 1824 GYTSSEFAADALLVFNDCQTFNEDDSEVGKAGHIMRRFP 1863
Db 1605 EYKLASEFIDDIELMFSNCFEYNPRINTSEAKAGTRLOAFF 1644
```

RESULT 23

```
US-09-839-479-69
; Sequence 69, Application US/09839479
; Publication No. US20020039779A1
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042002
; CURRENT APPLICATION NUMBER: US/09/839,479
; CURRENT FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 09/418,710
```

```
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 69
; LENGTH: 1673
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-839-479-69
```

Query Match 6.8%; Score 669; DB 9; Length 1673;

Best Local Similarity 19.7%; Pred. No. 1.7e-27;
Matches 303; Conservative 200; Mismatches 397; Indels 640; Gaps 43;

```
QY 644 RGRGRPPKVTITELNKTNRPLKLEAQETLNEEDKAKIAKSKKMKROKVORGECLTTI 703
Db 424 RRRGRPK-RIH--ISOEDN-----VANKQTLASYSKATKERDKLLKQEMKSL 470
QY 704 QGQARNKRQETKSLKHKEAKKKSXAEBKRGKTQKQKLEKVKREKKEKVKQKKEEYTK 763
Db 471 AFEKAKLKREKADAL---EAKKKEKEDKEK---KREELKKIVEERLKKKEERLKVVER 524
QY 764 AKPACAKADTLATORRLEERQKQMIILEMKKPTEDMCLTDHQPLPDFSRVPGITLPSGA 823
Db 525 EK-----EREKLEERKRYVEYKQWSKPREDMCECDLDELPEPTVK-TRLPPEI 573
QY 824 FSDCLTIVFLHSPGKVLGPDPAKDVP---SLGVLOEGLLCCQDGLSGEVDOLLVRLKAA 880
Db 574 FGDALMWLEFLNAGEL--FDLQDEFPDGVTVLESEALV-GNDSEGPLCBLLFFFLTAI 630
QY 881 LHDPFPSPYCSQLKILGEKYSEIPLTRDN-----VSEILRCFLMAYGV-- 923
Db 631 P-----QAIAESEEEVAKEQLTDADTKGCSLSKSLDLDSCITLSEILRLHILASGADV 681
QY 924 -----XPALCDRLTQPPQAOPPOQAAPPOQAQAAVLAPFVHEL 955
Db 682 TSANAKRYQKRGGFATDDACMELRLSNFSLVKLSSTSVYDLTPGEMKIL-----HAL 737
QY 956 NGSTL-----IINEIDKYLESNSS--YRKNKWIVEGRLLRLKTVLAK----- 995
Db 738 CGKLLTLVSTRDFIEDYVDILRQAKQEFRELKAQHREKEREAAAARIKREKELKEQEQ 797
QY 996 -----RTGRSEVEMGRPE-----ECILGRRRRSSRIMEETSQMESEEEESTAA 1037
Db 798 KMKEQKELKXEDQERNSTADISIGEEREDFDTSIESKDEQKELDQDMFTEDDDPGSH 857
QY 1038 VPGRRGRDGEVDNATASSIPELERQ-----IEKLSKQLFRFKLLHSSQ 1082
Db 858 KRGRGRKRG-----QNGFKFTTQEQINCVCVTRRELLTADSEEAALKQHQRKEKELLEKQ 911
QY 1083 MLRAVS---LGQDRYRRRYWVLPYLAGIFVEG-----TEGNLVPBEEVIXKETDSLKVA 1133
Db 912 SAIACTNIFPLGRDRMTRRWIFFSLPGLFIEEDYSGLTEDHLLP----- 956
QY 1134 HASLNLPALFSMKMBELAGSNNTTASSPARARSRLPKTKFGFMQPRHFKSPVRGQDSEQPOAQ 1193
Db 957 -----RPSSFQNNVQSOD----- 969
QY 1194 LOPEAQLHVPAPQOPOLQLOLQSHKGFLEQEGSPLSLQSQSHLSQSFAFLSWLSQTSQSHS 1253
Db 970 -----PQV----- 978
QY 1254 SLLSSSVLTDPSSPGKLDPAQSOPPEPEDEAESSPDLOAFWFNISAQMPCNAAPTPL 1313
Db 979 PLMSESTNIDQGR-----DHSVQLP----- 1000
QY 1314 AVSEDOPTSPQOLASSKPMRPNANPCSPVQFSSTPLAGLAPKRRAGDGPMPQSTG 1373
Db 1314 AVSEDOPTSPQOLASSKPMRPNANPCSPVQFSSTPLAGLAPKRRAGDGPMPQSTG 1373
```

```
Db 1001 -----KPVHKPNR----- 1008
Qy 1374 LQPKRGRGPPSKPFQKMEQRYLTQTAQVPPKSCGMMWIPDPEDMLDAMLKALHPRGI 1433
Db 1009 -----WCFYSSCEQLDQLIEALNSRGH 1030
Qy 1434 REKALHKLHNRDPLQEVCLRPSADPIP-----EPRQLPAFQEGIMS--WSPKKTVE- 1485
Db 1031 RESALKETLLOEKSRICAQLARFSEKHFHSDKRPDSKPTYSRGRSSNAYDPSQCAEK 1090
Qy 1486 -TDLAVLQWVELEQVRIM-----SDLOIRGWTCPSPD-----STRE 1521
Db 1091 QLELRDLPLDIEDRIYQGTGALIKVTRDRIHRSALSGRYELLSEBENKENGIIKTVE 1150
Qy 1522 DLAYCEHLSDSQEDITWRGPGREGIAPQRKTNP-----DLAVMLAALEQNYKR 1572
Db 1151 DVEEMEIDEQTKVIVKDRLLGKTETPTSTVSTNASTPQSVSSVHYLAMALFQIEQGLER 1210
Qy 1573 RYLREPL-----WPHFVLEKALLSTPNGAPEGTTTISIEITPRIRWQTQRCRSA 1627
Db 1211 RFLKAPLDASDGRSYKTVLDR-----WRESLLSSASL 1243
Qy 1628 AHVCLCLGHLERSIAWESVNVKTVCLVCRKGNDFELLCDGDRGCHIYCHRPKWEAVP 1687
Db 1244 SOVFLHLSTLDRSVIWSKILNARCKICRKGDAENMVLCDCDRGHHTYCVRPKLIKIVP 1303
Qy 1688 EGDWFCPCPKQRCERLSFRQRPSEDSDEVEDSMGGEDDEVDGDEEGQSEBEEYEVE 1725
Db 1304 EGDWFCPCPKQRCERLSFRQRPSEDSDEVEDSMGGEDDEVDGDEEGQSEBEEYEVE 1363
Qy 1726 FSEGD-----GR---RRVLLKGR-----ESPAAGPRYSE----- 1752
Db 1364 QDEDDSEBEEVSLPKRGKRPQVLRKGLSSFSRSGQQQEPGRYPSPRSQQSTPKTT 1423
Qy 1753 -----EPLSPSKRR----- 1762
Db 1424 VSKTGRSLRKINSAPPTETKSLRIASRTHSHGHPLOADVFVELLSPRKRRGRKRSANN 1483
Qy 1763 -----LSMRNHS----- 1770
Db 1484 TPENSPFNPFVIATKSEQSRSVNIASKLSLQSESKRCRCRQSPSPVTLGRSS 1543
Qy 1771 -----DLTFCEIILMESHDAAWPFKEPVNPLVSGYRRIIKNPMDFSTMRELLRG 1823
Db 1544 GRQGVHLSAPELLVWELRHDSWPLKLVSKIQVPDYVDIHKPIALNIIREKVNC 1603
Qy 1824 GYTSSEFAADALLYPDNCOTFNEDDSEVGKAGHIMRRFF 1863
Db 1604 EYKLASEFIDDIELMFSNCFEYNPRNTSEAKAGTQLQAFF 1643
```

RESULT 24

```
US-10-376-537-70
; Sequence 70, Application US/10376537
; Publication No. US20030224405A1
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042001
; CURRENT APPLICATION NUMBER: US/10/376,537
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US/09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 70
; LENGTH: 1673
```

```
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-376-537-70

Query Match
Best Local Similarity 19.7%; Score 669; DB 15; Length 1673;
Matches 303; Conservative 200; Mismatches 397; Indels 640; Gaps 43;

Qy 644 RGRGRPPKVTITELLNKTNNRPLKLEAQTLEEDKAKIAKSKKQKQVQRGELTTI 703
Db 424 RRRGRPPK-RIH--ISQEDN-----VANKQTLASYSRKATKDKRLKQEMKSL 470
Qy 704 QGQARNKQKQTKSLKKEAKKKAEBKKTQKELKEKVKEKKEKVKMKEEYTK 763
Db 471 AFEKAKLREKADAL---EAKKKEKDKK---KREELKTIIVBERLKKKEKELKVER 524
Qy 764 AKPACKADKTLATQRRLEEROKQOMILEEMKPTEDMCLTDHOPLPDSRVPGLTPSGA 823
Db 525 EK-----EREKLEERKKVYKQWSPKPNEDMCCDLKELPEFPVK-TLPPPEI 573
Qy 824 FSDCLTIVFHLSPKVLGFDPKADVP---SLGVLQEGLLCGDSLGEVQDLLVRLKXA 880
Db 574 FGDALMWLEFLNAFGEL--FDLQDEFDPGVTVLELEALV-GNDSEGPLCELLPFLTAI 630
Qy 881 LHDPGPPSYCQSLKILGKGVSEIPLTRDN-----VSEILRCFLMAYGV-- 923
Db 631 F-----QAIABEEBEVAKELQTDADTKGCSLSDLDSDCTLSEILRLHILASGADV 681
Qy 924 -----XPALCDRLRTQPFQAPQPPQQAQAAVLAPFVHEL 955
Db 682 TSANAKRYQKRGDFDATTACMELRLSNPLVKLSSTSVYDLTPGEMKIL-----HAL 737
Qy 956 NGSTL-----IINEIDKTLSESS--YRKNKWIIVEGRRLRLKTVLAK---- 995
Db 738 CGKLLTVLSTDFIEDYVDILRQAKQEPRELKABQHRKEREAAARINKRKEKLKEQEQ 797
Qy 996 -----RTGRSEVEMGRPE-----ECLGRRRSRIMEETSMEEBEESESTAA 1037
Db 798 KMKQKELKDEQERNSTADISIGEEREDPDTSTESKDTQKELDQDMFTEDEDDPGSH 857
Qy 1038 VPGRRGRDGEVDATASSIPELERQ-----IEKLSKRQLFRKKLLHSSQ 1082
Db 858 KRGRGRKRG-----QNGFKFTQEQINCVTRELLTADEEALKQEHQKELKLEKIQ 911
Qy 1083 MLRAVS-----LGQDRYRRRYWVLPYLAGIFVEG-----TEGNLVPPEVIKKTETSLKVA 1133
Db 912 SAIACNIFPLGRDRMYRYWIFPLPLGLFIEDYSGLTEDHLLP----- 956
Qy 1134 HASLNPAFLFSKMKELAGSNTTASSPARARSPLKTKPGFMQPRHFKSPVRGQDSQPOAQ 1193
Db 957 -----RPSSFQNNVQSOD----- 969
Qy 1194 LQPEAQLHVPAQPOPOLQLQSHKGFLEQSGPLSLGQSQHDLSQSASFSLWSLQSHS 1253
Db 970 -----PQV-----STKTGE 978
Qy 1254 SLLSSSVLTPDSSPGKLDPAQPPPEPEPEDEASSPDLOAFWFNISAQMPNAAPTPL 1313
Db 979 PLMSESTNIDQGR-----DHSVQLP----- 1000
Qy 1314 AVSEDOPTPSQQLASSKPMNRPSAANPCSPVQFSSTPLAGLAKRRAGDGEEMPQSPG 1373
Db 1001 -----KPVHKPNR----- 1008
Qy 1374 LQPKRGRGPPSKPFQKMEQRYLTQTAQVPPKSCGMMWIPDPEDMLDAMLKALHPRGI 1433
Db 1009 -----WCFYSSCEQLDQLIEALNSRGH 1030
Qy 1434 REKALHKLHNRDPLQEVCLRPSADPIP-----EPRQLPAFQEGIMS--WSPKKTVE- 1485
Db 1031 RESALKETLLOEKSRICAQLARFSEKHFHSDKRPDSKPTYSRGRSSNAYDPSQCAEK 1090
Qy 1486 -TDLAVLQWVELEQVRIM-----SDLOIRGWTCPSPD-----STRE 1521
```

Db 1091 QLELRDLRDLIEDIRIYQGTILGAIKVTDRHWRSALESRGYELLSEENKENGIIKTVE 1150
Qy 1522 DLAYCEHLSDSQEDITWRGPGREGIAPQKTNP-----LDLAVMLAALEONVKR 1572
Db 1151 DVEEMEIDQTKVIVKDRLLGIKTETPTSTVSTNASTPOSVSVVHYLAMALFOEQGLER 1210
Qy 1573 RYLRPL-----WPTHEVLEKALLSTNGAPEGTTTISYEITPRIRIWRQTLQRCRSA 1627
Db 1211 RFLKAPLDASGSRKYVLDL-----WRESLLSSASL 1243
Qy 1628 AHVCLCGLHLERSIAWESVNVKVTCLVCRKGNDEFLLLCDGCDRGCHYCHRPKMEAVP 1687
Db 1244 SQVFLHLSTLDRSVIWSKSIILNARCKICRKGDAENMVLCDCDGRGHHTYCVRPKLIKVP 1303
Qy 1688 EGDHFCIVCLAQ-----VEGEFTQKPGFPKRGOKRSGYSLN 1725
Db 1304 EGMFCPECRPKQRCRLSRQPSLESDVEDPSMGGEDVDGDEBEGQSEEEVE 1363
Qy 1726 FSEG-----GR---RRVLKGR-----ESPAAGPRYSE----- 1752
Db 1364 QEDDSQEEBVSPLKRGKPOVRLPVKTRGKLSFSSRGQQQBPGRYPSPRSQOSTPKTT 1423
Qy 1753 -----BRLSPSKERR----- 1762
Db 1424 VSSKTRSLRKINSAPPTETKSLRIASRSTRHSHGHPLOADVVELLSPPRRKRGKRSANN 1483
Qy 1763 -----LSMRNHS----- 1770
Db 1484 TPENSPNPNFRVIATKSSESRSVNIASKLSQESKRRCKRKQSPSPVTLGRRSS 1543
Qy 1771 -----DLTFCILMEMESHDAWPPEKVPNPLVSGYRIIKNPMDFSTMBERLLRG 1823
Db 1544 GRQGVHLSAFELQVLVLRHDSWPFLKLVSKIQVPDYDIKKPITALNIIREKVNKC 1603
Qy 1824 GYTSEEPFADALIVFONCOTFNEDDSEVGKAGHIMRRFF 1863
Db 1604 EYKLASEFIDIELMFSCFEYNPRINTSEAKAGTLQAFF 1643

RESULT 25

US-10-702-148-69
; Sequence 69, Application US/10702148
; Publication No. US20040063145A1
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042002
; CURRENT APPLICATION NUMBER: US/10/702,148
; CURRENT FILING DATE: 2003-11-05
; PRIOR APPLICATION NUMBER: PRIOR APPLICATION NUMBER: US/09/839,479
; PRIOR FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ IDS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 69
; LENGTH: 1673
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-702-148-69

Query Match 6.8%; Score 669; DB 15; Length 1673;
Best Local Similarity 19.7%; Pred. No. 1.7e-27;
Matches 303; Conservative 200; Mismatches 397; Indels 640; Gaps 43;
Qy 644 RGRGPPKVKITELLNKTNDNPLKLEAQETLNEEDKAKIAKSKKQKQVQRGECITTI 703

Db 424 RRRGRPFK-RIH--ISQEDN-----VANQTLASYRSKATKERRDKLLQBEEMKSL 470
Qy 704 QGOARNRKQBTKSLKHKEAKKSAKEKTKTQEKLKKEVKREKKEKVKQMKKEEVTK 763
Db 471 AFEKAKLREKADAL---EAKKKEKEDKEK---KREELKKIIVEERLKKKEKERLKYER 524
Qy 764 AKPACKADKTLATORRLEERQKQOMILEEMKPTEDMCLTDHQLPLDPSFVPGTLPLSGA 823
Db 525 EK-----EREKUREKKRYFYKQSKPRDECDLDELPEPTPVK--TRLPPPI 573
Qy 824 FSDCLTIVEFLHSFGKVLGDFDPAKDPV---SLGVLQBLGCLQCGSLGSEVQDVLRLKAA 880
Db 574 FGDALMWLEFLNAGEL--FDLQDEFPGDVTLEVEALV-GNDSGEGPLCELLFFLTAI 630
Qy 881 LHDPGFFSYCQSLKILGEKYSIEPLTRDN-----VSEILRCLFLMAYGV-- 923
Db 631 F-----QAIABEEBEVAKEQLTDADTKGCSLKSILDLDSCTLSEILRLHILASGADV 681
Qy 924 -----XPALCDRLRTQPFQAPPOQPOQKAALVAPFVHEL 955
Db 682 TSANAKRYQKRGDFDADDAEMELRLSNPSLVKLSSTSVYDLTPGEMKIL---HAL 737
Qy 956 NGSTL-----IINEIDKTLESMS--YRKQKWIIVEGRRLRLKTVLAK---- 995
Db 738 CGKLLTLVSTDFIEDYVDILROAKQBFRELKASQHKREREAAAARIRKKEKLEQEQ 797
Qy 996 -----RTGRSEVMGRPE-----ECUGRRSSRIMBEETSGMEBEEBESIAA 1037
Db 798 KMKEKQEKLEDEQORNSTADISIGEEREDPDTIESKDTQEKELDQDMFTEDEDDPGSH 857
Qy 1038 VPERGERDGEVDATASSIPELREQ-----LEKLSKROLFRFKLLHSSQ 1082
Db 858 KRGRGRKRG-----QNGFKEFTQEQINCVTRELLTADBEELAKQEHQKKELEKLQ 911
Qy 1083 MLRAVS-----IGQDQYRRRYWVLPYLAGIFVEG-----TEGNLVEEVIKKTDSLKVA 1133
Db 912 SAIACTNIFPLGRDRMYRYWIFPLPGLFIEDYSGLTEDHLLP----- 956
Qy 1134 HASLNPALFSKMKELAGSNTTASSPARARSRLTKPGFMQPRHFKSPVRQDSEQPOAQ 1193
Db 957 -----RPSSFQNNVQSD----- 969
Qy 1194 LQPEAQLHVPAQPOPQLQLOLQSHKGFLEQEGSPLSLGSOHDLSSQAFSLWSLQTSHS 1253
Db 970 -----PQV-----STKTGE 978
Qy 1254 SLASSSVLTPDSSPGKLDPAPOPPPEPEDEABESSPDLOAFWFENISAQMFCAAPTPL 1313
Db 979 PLMSESTSNIDQGR-----DHSVQLP----- 1000
Qy 1314 AVSEQDTPSPQQLASSKPMNRPSAANPCSPVQSFSTPLAGLAPKRRAGDGPMPQSPG 1373
Db 1001 -----KPVHKPNR----- 1008
Qy 1374 LGQPKRGRPPSKFFKQMEQRYLTQLTAAQVPPPEMCSGMMWIPDPEMLDAMLKALHPRGI 1433
Db 1009 -----WCFYSCEQLDQLEALNSRGH 1030
Qy 1434 REKALHKLHKNHRDLQEVCLRPSADPIF-----EPRLPAFAQGIGMS--WSPKEKTYE- 1485
Db 1031 RESALKETLLOEKSRICAQLARFSEEFHPSDKKQPSKPTYSGRGSNADVPQMCNAEK 1090
Qy 1486 -TDLAVLQWVEELEQVRVIM-----SDLOIRGWTCPSPD-----STRE 1521
Db 1091 QLELRDLRDLIEDIRIYQGTILGAIKVTDRHWRSALESRGYELLSEENKENGIIKTVE 1150
Qy 1522 DLAYCEHLSDSQEDITWRGPGREGIAPQKTNP-----LDLAVMLAALEONVKR 1572
Db 1151 DVEEMEIDQTKVIVKDRLLGIKTETPTSTVSTNASTPOSVSVVHYLAMALFOEQGLER 1210
Qy 1573 RYLRPL-----WPTHEVLEKALLSTNGAPEGTTTISYEITPRIRIWRQTLQRCRSA 1627


```
Db 903 GTDRNHNRYWLFSDVPLGFIE-----KGWVHDSID----- 933
Qy 1149 AGSNTTASSPARARSRLKTRPGFMQPRHFKS-PVRGODSEPOQAQLOPEAQLHVPAPQ 1207
Db 934 -----YRFNHCKDHTVSGCEDYCPRSK----- 956
Qy 1208 PQLQLQSHKGFLQEGSPISLGOSQHDLSQSAFLSQTQSHSSILLSVLTDPSSP 1267
Db 957 -----KANLGN-----ASMTQHGTAATEVAVETTPKQ 985
Qy 1268 GKLDAPSPQPEPEDEAESPDLOAFWNISQMPCNAAPTPLAVSEDPQTPSPQOL 1327
Db 986 G-----QNLWFLCDSQ----- 996
Qy 1328 ASSKPMNRPAAANPCSPVQFSSTPLAGLAPKRRAGDPGEMPQSPGTGLQPKRGPPSKF 1387
Db 997 ----- 996
Qy 1388 FKQMEQRYLTQLTQAQVPVPEMCSGWWIIPDPEMLDAMLKALHPRGIREKALHKLHKLK-HR 1446
Db 997 -----KELDELLNCLHPQGIRESQLEKREKRYQ 1025
Qy 1447 DFLQEVCLRPADPIFEPRQPLPAFQEGIMSNPKB-----KTYETDLAVLQWVEELRQ 1500
Db 1026 DIIHSLHL--ARKFNGLKSCDGNQE--LLNFLRSLDIEVATRLQKGLGYVEETSEFEAR 1082
Qy 1501 VIMSDLOIRGWTCSPDSTREDLAYCEHLSDSQEDITWRGGRGLAPQRTTNPLDLAV 1560
Db 1083 VIS-----LEKLKDFGECV----- 1096
Qy 1561 MRLAALQONKRRVIREPLWP-----THEVWLEKALLSTPNGAPEGTTTEI 1606
Db 1097 ---TALQASVLIKFLQGMAPKQKRKLQSEDSAKTEVDEKKV----- 1140
Qy 1607 SYEITPRIIRIWRQTLQRCSAAHVCLCLGHLERSIAMEKSNVYKTVCLVCRKGDNDFLL 1666
Db 1141 EAKVASALEKWTAREAQTSRHHVLLGLMDACIKWMSAENARCKVCPKKGDDKLIL 1200
Qy 1667 CDGCDRGCHYCHRPKMAVPEGDFWCTVCLAQQVEGEFTQKGFPPKRGQKRGYSINF 1726
Db 1201 CDECNKAPHLFLRPALYEVDPGEWQCPAC-----OPATARNRSGRNYTEESA 1249
Qy 1727 SE-----GDGRRRRVLLKGRES---PAA--GPR-----YS 1751
Db 1250 SEDSEDESDESEEEEEEDYEVAGLRRLRPRTIRKGHSVIPPAARSRRPGKKPHS 1309
Qy 1752 EERLSPSK-----RRRLSMRNHSDLTFCIEILMEMESHDAWPFXPFPVNPRL 1799
Db 1310 TRRSOPKAPPVDAEYDELVLQTKRSSRRRQSLQLEKCEBILHKIVKYRFSWPFREPVTRDE 1369
Qy 1800 VSGYRIIKNPDMFSTWRELLRGGYTSSEFEADALLVFNQCOTFNEDDSEV 1852
Db 1370 AEDYDVITHPMDFTQVQNKSCGSYSVQBFETMDKQVFTNAEYVINCGRSHV 1422
```

RESULT 28

```
US-10-376-537-69
; Sequence 69, Application US/10376537
; Publication No. US20030224405A1
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; CURRENT APPLICATION NUMBER: US/10/376.537
; CURRENT FILING DATE: 2003-03-28
; PRIOR APPLICATION NUMBER: US/09/418, 710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
;
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 69
; LENGTH: 1525
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-376-537-69

Query Match 5.3%; Score 521.5; DB 15; Length 1525;
Best Local Similarity 19.6%; Pred. No. 1.6e-13;
Matches 281; Conservative 185; Mismatches 444; Indels 523; Gaps 48;

Qy 565 EVIKYLVGNLVHSVRREHFSFSPRMP-----VGDFEERDTPGELQWQLSAEELPS 616
Db 368 EMWQMSFNKLHT-----NPHIPKGGPPAKPG---KHSDFP-----LXAKGRSK 409
Qy 617 RI---QAITKRGGRPR--NTEKAKTKVEPKVGRGRPPKVKITEL-----LNKTDNR 664
Db 410 GILNGQKSTGNSPKKGLKTPKTKMQWTLDDMAKGTQ---KMTRAPRNSGGTPTRTSSK 466
Qy 665 PLKKLEA-----QETLNEEDKAKIAKSKKQKQVQGECLTTIQOARNKRQKET 715
Db 467 PHGHLPPAALHLIAIYYKENKDRDKRSALSCVISTKARTLLSSEDRARLPBELRSLVKRY 526
Qy 716 KSLKHKE--AKKKSXAEEKGKTQKELKSKVKREKKEKVKMKEKEEVTRAKAPACKADKT 773
Db 527 ELLEHKRWASMBEQRKEYLKKREELKKLKEKAKER---REKE----- 569
Qy 774 LATQRRLERQKQOMILEEMKPTEDMCLTDHQPDPFSRVPGLT-LPSGAFSDCLTIVE 832
Db 570 ----MLERLEKQ-----KRYEDQELTG-KNLPAFLVDTPEGLPNTLFGDVAMVVE 615
Qy 833 FLHSFGKVLGDPKADVPISLGVLOEGLLCOGDSLGEVQDILLVRLKKAALHDPGFPSCQS 892
Db 616 FLSCYSGLLLPDAQYPIYAVS-LMEALADKGGFL-YLNRVLVILLQLODEIAEDYGE- 672
Qy 893 LKILGEKVSIPILTRDNVNSBILRCFLMAYG-----VXPALCD 929
Db 673 ---LGMKLSIPLTLHSVSELVRLCLRRSDVQSESGSDTDDNKDSAAFEDNEVQDEFLE 729
Qy 930 RLRTQPPQAQPOOQAAVLAPFVHLENGSTLIINEIDKTLSESSSYRNKWKVIGRULRL 989
Db 730 KLETSEFPFELTSEBKQILTALCHRI----LMTYSVQDHMETROQMSAELW--KERLAVL 783
Qy 990 K----TVLAKRTGRSEVEM-----GRPECLGRRRSRIMEETSGMBEEEEESIAAVPG 1040
Db 784 KEENDKRAEKQKREKEMKENGKVENGLGKTRDKKRIIVKFFEPQVDTEADDMISAVKS 843
Qy 1041 RR-----GRDGEVDATASSTPELERQIEKLSKR-----QLFPRKKLLHSSQMLRAVSL 1089
Db 844 RLLAIQAKKEREIOEREMKV-KLERQAESEIRIKHKAABAKAFQEGIAKAKLVRRTPI 902
Qy 1090 GQDRIYRRRYWL-PYLAGIFVEGTGELNVPVEVIKKTSLKVAHAASLNPALFSMKMWEL 1148
Db 903 GTDRNHNRYWLFSDVPLGFIE-----KGWVHDSID----- 933
Qy 1149 AGSNTTASSPARARSRLKTRPGFMQPRHFKS-PVRGODSEPOQAQLOPEAQLHVPAPQ 1207
Db 934 -----YRFNHCKDHTVSGCEDYCPRSK----- 956
Qy 1208 PQLQLQSHKGFLQEGSPISLGOSQHDLSQSAFLSQTQSHSSILLSVLTDPSSP 1267
Db 957 -----KANLGN-----ASMTQHGTAATEVAVETTPKQ 985
Qy 1268 GKLDAPSPQPEPEDEAESPDLOAFWNISQMPCNAAPTPLAVSEDPQTPSPQOL 1327
Db 986 G-----QNLWFLCDSQ----- 996
Qy 1328 ASSKPMNRPAAANPCSPVQFSSTPLAGLAPKRRAGDPGEMPQSPGTGLQPKRGPPSKF 1387
Db 997 ----- 996
Qy 1388 FKQMEQRYLTQLTQAQVPVPEMCSGWWIIPDPEMLDAMLKALHPRGIREKALHKLHKLK-HR 1446
```

[illegible]

```

RESULT 29
US-10-702-148-68
; Sequence 68, Application US/10702148
; Publication No. US20040063145A1
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042002
; CURRENT APPLICATION NUMBER: US/10/702,148
; CURRENT FILING DATE: 2003-11-05
; PRIOR APPLICATION NUMBER: PRIOR APPLICATION NUMBER: US/09/839,479
; PRIOR FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 68
; LENGTH: 1525
; TYPE: PRF
; ORGANISM: Homo sapiens
US-10-702-148-68

Query Match      5.3%; Score 521.5; DB 15; Length 1525;
Best Local Similarity 19.6%; Pred.No. 1.6e-19;
Matches 281; Conservative 185; Mismatches 444; Indels 523; Gaps 48;

Qy    565 EVIKYLRNLVHSVRREHFSRPM-----VGDFEERTPEGLQWQLSAEEIPS 616
       ||:||::|||:||||:||||:||||:||||:||||:||||:||||:||||:
Db    368 EMMKMSFNKLHT-----NPHIPKKGGPPAKPKG---KHSDKP-----LKAGRSK 409
Ov    617 RI----QAITGRGRPR---NTEAKTKVEPVKRGRRGPVKVTTEL-----LNKTDNR 654
```

Db	410	GILNGKSTGNSKSPKKGLKTPKTKMOMTLDMWAGTQ---KMTAPRNSGCTPRTSSK	466
Qy	665	PLKKLEA-----QETLNEEDKAKIAKSKKKMKQKQVORGECLETTIOGAARNEKKEQT	715
Db	467	PKHCLPAAHLHLLIAYKENKQREDKSALSVCVLSIKSTARLLSSDDRARLEPRLSVQKRY	526
Qy	716	KSLLKHKE--AKKXSXAEEKGKTKQEKLEKVKVREKKEKVKMKKEBEVTKAKPACKADKT	773
Db	527	ELLEHKRWASMSSEQRKEYLKKKRBELKKLKEKAKER-----REKE-----	569
Qy	774	LATQRLREERQKQOMILEEMKKPTEDMCITDHOPLDPDRSRVPLGT-LPSGAFSDCLTIVE	832
Db	570	-----MLERLEKQ-----KRYEDQELTG-KNLPAFLVDTPEGLPNTLFGDVAMVVE	615
Qy	833	FLHSFGKVLGFDPAKDVPSGLVQLQEGLLCOGDSLGSVQQLLVRLKKAALHDPGFSPYCQS	892
Db	616	FLSCYSGLLLPDAQYPIITAVS-LMEALADKGGFL-YLNRVIVLLOTLQLODETAEDYGE-	672
Qy	893	LKTLGKVSSEIPUTRDNVSEILRCFLMAYG-----VXPALCD	929
Db	673	---LGMKLSBEIPLTHSVSELVRCLRRSDVQSESGSDTDNMDKSAAFEDNEVQDEFLE	729
Qy	930	RLTQTFQAPQPOOKAAVLAPFVHELNGSTLIINEIDKTLSESSYRKXNKWIVEGBLRRL	989
Db	730	KLETSEFFELTSEKLIULTALCHRI---LMTYSVQDHMETRQQMSAELW--KERLAVL	783
Qy	990	K---TVLAKRTGRSEVEM---GRPEECIGRRSSRIWEETSGMEEEEBEESIAAVPG	1040
Db	784	KEENDKRAEKQKRWKEMAKENGKVENGLKTRDKRKIRIVKEPQVDTEAEADMI SAVKS	843
Qy	1041	RR-----GRDGEVDATASSIPELERQIEKLSKR-----QLPFRKLLHSSQMLRAVSL	1089
Db	844	RLLLAIQAKKEREIQEREMKV-KLERQAEERIRKHAAAEAKAFQSGIAKAKLVMRRTPI	902
Qy	1090	GQRYRRRYVWL--PYLAGIFVEGTEGNLVEEVIKETDSLKVAHAASLNPALFSMKMEL	1148
Db	903	GTDNRHNRYLFDVEPGLFIE-----KGWVHDSID-----	933
Qy	1149	AGSNTTASSPARARSPLTKPGFMQPRHFKS--PVRGQDSEQPOAQLQPEAQLHVPAPQ	1207
Db	934	-----YRNNHCKDHTVSGCEDYCPRSK-----	956
Qy	1208	PQLQLQLQSHKGFLEQEGSPLSGQSHDLQSASFILWSLQTSQSHSSLLSSSVLTDPDSSP	1267
Db	957	-----KANLQKN-----ASMNTOHGATATEVAVETITPKQ	985
Qy	1268	GKLDPAPOPPPEPEDEAESSDLAQAFWNISQAQPCNAAPTTPPLAVSEDQTPSPQQL	1327
Db	986	G-----QNLWFLCDSQ-----	996
Qy	1328	ASSKPMNRPSAANPCSPVQFSSTPLAGLPKRAAGDPGEMPOSPTGLGQPKRGRPPSKF	1387
Db	997	-----	996
Qy	1388	FKQWEQRYLTQLTAQVPVPEMCSGMWIIPDEMLDAMLKALHPGRIREKALHKHLNK-HR	1446
Db	997	-----KELDELLNCLHPQIGIRESOLKERLEKRYQ	1025
Qy	1447	DFLOEVCRLPSADPIFEPQLPAFOGINSWSPE-----KTYETDVLAVLOWVEELEQR	1500
Db	1026	DIITHSLHL--ARKPNLGLSKDCGNQE--LLNFLRSLDLIEVATRLQKGLGYVEETSEFEAR	1082
Qy	1501	VIMSDLOIRGWTCFSPDSITREDLAYCEHLSDSODEITWRCPGREGPLAQPKRTNPLDLAV	1560
Db	1083	VIS-----LEKLQDFGECV-----	1096
Qy	1561	MRLAALAEONKRRYLREPLWP-----THEVVLEKALLSTPNGAPEGTTTEI	1606
Db	1097	---IALQASVIKFLQGFQWAPKQKRKLOSEDSAKTEESVDEEKKWVE-----	1140
Qy	1607	SYBITPRIRIWRQTLQRCSAAHVCLCLGHLESI AWKSVNKTCLVCRKGDNDLFLL	1666

```
Db 1141 EAKVASALEKWKTAIREAQTSRMHVLLGMLDACLKWDMSAENARCKVCPKKGGDKLIL 1200
Qy 1667 CDGDRGCHYCHRPKMEAVPEGDFCTVCLAAQVEGEFTQKPGPPKRGQKRGYSYNF 1726
Db 1201 CDECNKAPHLFLPALYVEPDGEMQCPAC-----QPATARRNSGRNYTESSA 1249
Qy 1727 SE-----CDGRRRRVLLKGRES---PAA--GPR-----YS 1751
Db 1250 SEDSEDESEDEEBEEDYEVAGRLRPRTKIRGKHSVIPPAARSRRRGKKPHS 1309
Qy 1752 EERLSPSK-----RRRLSMRNHSDLTFCETILMEMESHDAAPFPXFPVNPRL 1799
Db 1310 TRRSQKAPPVDAEDELVLQTKSSRRQSLQCEILHKKVYKRFSPFPREPVTRDE 1369
Qy 1800 VSGYRIIKNMDPSTMRERLLRGYTSSEFAADALLVFNQCOTFNEDDSEV 1852
Db 1370 AEDYDVITHPMDFTQVQNKSCGSYSRVQBFLLDMKQVFTNAEYVNCRGSHV 1422

RESULT 30
US-09-839-479-27
; Sequence 27, Application US/09839479
; Publication No. US2002003979A1
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042002
; CURRENT APPLICATION NUMBER: US/09/839,479
; CURRENT FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 27
; LENGTH: 1527
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-839-479-27

Query Match 5.3%; Score 521.5; DB 9; Length 1527;
Best Local Similarity 19.4%; Pred. No. 1.6e-19;
Matches 264; Conservative 179; Mismatches 440; Indels 479; Gaps 42;

Qy 619 QAITGKGRPR--NTEKAKTKEVPKVGKRGPRPKVKITEL-----LNKTDNRPLKGL 669
Db 415 QKSTGNSKSPKKGUKTPTKWKQMTLDMAGTQ---KMTAPRNSGGTPTRTSKPHKL 471
Qy 670 EA-----QETLINEEDKAKIAKSKKMKQKQVQGECLTTIOQQAARKKQKQETKSLKH 720
Db 472 PPAALHLIAYKENKOREDEKESALSVCVISTARLLSSEDRARLPEELSLVQKRYELLEH 531
Qy 721 KE--AKKXSAEKKGKTKQEKLEKVKREKKEKVKMKKEEVTAKPACAKADKTLATQR 778
Db 532 KKRWASMEEQREKYLKKREELKKLEKAKER-----REKE-----569
Qy 779 RLEERQOQMILEMKKPTEDMCLTDHQPDPFSRVPLGT--LPSGAFSDCLTIVEFLHSF 837
Db 570 MLELEKQ-----KRYEQQELTG--KNLFAFLVDTPEGLPNTFLFGDVAWVVEFLSCY 620
Qy 838 GKVLGFPAKDVPSLGLVQEGLLQCGDSLGVQDILLVLLKAAALHDPGFPSYCSOLKILG 897
Db 621 SGLLLPDAQYITAVS--LMEALSADKGGFLYNRVLLQTLQDETAEDYGE-----LG 675
Qy 898 EKVSEIPLTRDNVSEILRCFLMAYG-----VXPALCDRLRTO 934
Db 676 MKLSEIPLTHSVSELVRLCLRRSDVQESRGSDDTDNKKDSAAAFEDNEVDQEFLEKLETS 735
```

```
Qy 935 PFOAQPPQOKAAVLAFPMVHELNGSTLIINEIDKTTLESMSYRKKNKWIVEGRLRLK---- 990
Db 736 EFPFLTSEEKLIITLALCHRI----LMTYSVQDHMETRQOMSABLW--KERLAVLKEEND 789
Qy 991 TVLAKRTGRSEVEM-----GRPECLGRRSSRIEETSGMEEBEESIAAIVPGR-- 1042
Db 790 KKRAEKQKCKMEAKNKENGKVENGLKTRKRIKVFPEQVDTAEADMSIAVKSRRLLA 849
Qy 1043 --GRRDEVDATASSIPELEROIEKLSKR-----QLFFRKLLHSSOMLRAVSLGQRY 1094
Db 850 IQAKKEREIOBERMKV--KLERQAEERIRKHAAAEKAFQEGIAKAKLWMRTPIGTDRN 908
Qy 1095 RRRYVWL--PYLAGIFVSEGTGNLVPEEVIKKETDSLKVAHASLNPALFSMKMELAGSNT 1153
Db 909 HNRVWLSDEVPGLFIE-----KGVVHDSID----- 934
Qy 1154 TASSPARARSPLKTKFQMQRHFKS--PVRGDSQEQPAQLQPEAQLHVPAQPQLQL 1212
Db 935 -----YRFNHCKDHTVSGDEDYCPRSK----- 957
Qy 1213 QLQSHKGFLBQEGSPLSGQSQHDLSQSAFLSWLSQTSQSHSSLLSSSVLTDPDSSPKLDP 1272
Db 958 -----KANLGKN-----ASMTQHGTATEVAVETTPKQG----- 987
Qy 1273 APSQPEEPPEDEAESPDQLQAFWFNISAQMPCNAAPTPPLAVSEDOPTSPQOLASSKP 1332
Db 988 -----QNLWFLCDSQ----- 997
Qy 1333 MNRPSAANPCSPVQFSTPLAGLAPKRRAGDPGEMPOSPTGLGQPKRGRPPSKFFKOME 1392
Db 998 ----- 997
Qy 1393 QRYLTQLTQAQVPPPMCSGWWIIPDPEMLDAMLKALHPRGIREKALHKHLNK--HRDFLOE 1451
Db 998 -----KELDELLNCLHPQGIRESQLKERLEKRYQDIHS 1031
Qy 1452 VCLRPSADPIPEPQPLPAFOEGIMSWSPKE-----KTYETDLAVLOWVELEORVMSD 1505
Db 1032 IHL--ARKPNLGLKSCDGNQE--LLNFLRSDLIIEVATRLQKGLGYVETSEFEARVIS-- 1086
Qy 1506 LQIRGWTCPSPDSTREDLAYCEHLSQSDQITWRGPGREGIAPQRTNPLDLAVMLAA 1565
Db 1087 -----LEKLKDFGECV-----IA 1099
Qy 1566 LEQNVKRYLREPLWP-----THEVLEKALLSTNGAPEGTTTISYEIT 1611
Db 1100 LQASVIRKFLQGFNAPKQKRKLQSEDSAKTEEYDEEKQWYE-----EAKVA 1146
Qy 1612 PRIRWOTLQRCRSAAHVCLCLGHLERSIAWEKSVNKKVTCLVCRKGDNDFFLLDCGCD 1671
Db 1147 SALEKWKTAIREAQTTFSRMHVLLGMLDACIKWMSAENARCKVCPKGGEDDKLILCDECN 1206
Qy 1672 RGCHYCHRPKMEAVPEGDFCTVC-----LAQQVEGEFTQKPGPK 1713
Db 1207 KAPHLFLCLPALYVEPDGEMQCPACQATARRNSGRNYTEESASEDSEDESEEBE 1266
Qy 1714 RQKRGKSGSYLNFSSEGGRRRVLLKGRES---PAA--GPR-----YSEERLSPSK---- 1759
Db 1267 EEEEDYEV---AGLRLPRKTIKRGHSVIPPAAARSRRPGKKPHSTRSQKAPPVD 1323
Qy 1760 -----RRRLSMRNHSDLTFCETILMEMESHDAAPFPXFPVNPRLVSGYRIIKNP 1810
Db 1324 DAEVDELVLQTKSSRRQSLQCEILHKKVYKRFSPFPREPVTRDEADYDVITHP 1383
Qy 1811 MDFSTMRERLIRGGYTSSEFAADALLVFNQCOTFNEDDSEV 1852
Db 1384 MDTQVQNKCSGYSYSRVQBFLLDMKQVFTNAEYVNCRGSHV 1425
```

RESULT 31
US-10-376-537-27
; Sequence 27, Application US/10376537
; Publication No. US20030224405A1

Db 621 SGLLLPDAQYPTAVS-LMEALSADKGGFLYNLRVLVILLQTLQTLQDLQDAIEDYGBLG 679
Qy 898 EKVSEIPLTRDNVSEILRCFLMAYG-----VXPALCDRLRTQ 934
Db 680 MKLSEIPLTHSVSELVRLCLRRSDVQESBGSDDTDNKKDSAAFPEDNEVQDEFLEKLETS 739
Qy 935 PFOAQPOOKAAVLAFPPVHELNGSTLIINEIDKTLSSMSYRKWKWIVVEGRLLRLK----- 990
Db 740 EFFELTSEEKQILTALCHRI-----LMTYSVDHMETRQMSAELW--KERLAVLKEEND 793
Qy 991 TVLAKRTGRSVEV-----GRPEECJGRRRSRRIEETSGMEEBEESIAAIVGRR--- 1042
Db 794 KKRAEKQKRKEWAKNKENGKVENGLGKTRKRIKVPFQPDVTEAEDMISAVKSRLLA 853
Qy 1043 --GRDGEVDATASSIPELERQIEKLSK-----QLFFRKLLHSSQMLRAVSLGQDQRY 1094
Db 854 IQAKEREIQEREMKV-KLERQAEERIRKHAAAEKAFQSGIAKAKLVMRTPTIGTDNR 912
Qy 1095 RRRYVWL-PYLAGIFVEGTEGNLVPVEVKKETDSLVKVAHAASLNPALPFSMKMELAGSNT 1153
Db 913 HNRWLFSDVPGFLFIE-----KGWVHDSID----- 938
Qy 1154 TASSPARARSPLTKTKCFMQPRHFKS-PVRGQDSEQPQAOQLQPEAQLHVPAPQPOQL 1212
Db 939 -----YRFNHCKOHTVSGDEYDCPRSK----- 961
Qy 1213 QLQSHKGFLEQSGPLSQSHDLQSQAFLSWLSQTSQSHSSLSLSSSVLTPOSSPGKLD 1272
Db 962 -----KANLGKN-----ASMNTQHGTAETAVETTPKQ----- 991
Qy 1273 APSQPEPEDEAESPDQAFWFNIQAQMPCNAAPTPLPVLAVSEDQTPSPQOLASSKP 1332
Db 992 -----QNLWFLCDSQ----- 1001
Qy 1333 MNRPSAANPCSPVQFSSTPLAGLAPKRAAGDPGEMQSPQTLGQPKRGRPPSKFFKQME 1392
Db 1002 ----- 1001
Qy 1393 QRYLTQLTAQVPVPMCSGWMMIPDPEMLDAMLKALHPGRIKREKALHKLHNK-HRDFLOE 1451
Db 1002 -----KELDELLNCLHPQIGRESOLKREKRVQDIIHS 1035
Qy 1452 VCLRSADPIFEPRQLPAFQSGIMWSPK-----KTYETDLAVLOWVELEQRVMSD 1505
Db 1036 IHL--ARKPNLGKSCDGNQ--LLNFLRSLDIEVATRLQKGLGVEETSEFARVIS-- 1090
Qy 1506 LQIRGWTCPSPDSTREDLAYCEHLSDSQSDITWRCPGREGLAPQRKTNPLDLAVMLAA 1565
Db 1091 -----LEKLKDFGECV-----IA 1103
Qy 1566 LEQNVKRRYLREPLWP-----THEVVLEKALLSTPNGAPEGTTTETISYEIT 1611
Db 1104 LOASVKKFLOGFMAPKQKRKLQSEDSAKTEEVDEEKWV-----EAKVA 1150
Qy 1612 PRIRIWRQTLQRCSAAHVCLGHLERSIAWEKSVNVTCLVCRKGNDNDFLLDCGCD 1671
Db 1151 SALEKWKTAIREAQTFSSRMVLLGLMDACIKWDSNAENARCKVCPKGGEDKLLCDBCN 1210
Qy 1672 RGCHYICHRPKWEAVPEGDWCTVC-----LAQVGEFTQKPGPK 1713
Db 1211 KAFHLFCLRLPALYEPDGEWQCPACQATARRNSGRNRYTESADESDESEDEEE 1270
Qy 1714 RGQKKSYSLNFSGDCGRRRLVLLKGRES--PAA--GPR-----YSEELSPSK---- 1759
Db 1271 EEEEEEDYEV--AGLRLPRKTIIRGHSHVIPPAAARGRPPGKPHSTRSQKAPVD 1327
Qy 1760 -----RRRLSMRNHSDLTFCETILMEMESHDAAPFEPNPLRVSGYRRIIKNP 1810
Db 1328 DAEVDELVLQTKRSRRSLELQKCEEILHKIVKVFWSPPREPVRDEAEDYDVITHP 1387
Qy 1811 MDFSTMRELLRGVTSSEFAADALLVDFNCQTNEDDSEV 1852
Db 1388 MDFQTVQNKCSGYSRVSQVEFLTDKMQVFTNAEVNCRGSHV 1429

RESULT 34

US-10-376-537-29
; Sequence 29, Application US/10376537
; Publication No. US20030224405A1
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042001
; CURRENT APPLICATION NUMBER: US/10376.537
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US/09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 29
; LENGTH: 1531
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-376-537-29

Query Match 5.24; Score 512.5; DB 15; Length 1531;
Best Local Similarity 19.24; Pred. No. Se-19;

Matches 262; Conservative 179; Mismatches 446; Indels 475; Gaps 41;

Qy 619 QAITGKGRPR--NTEKATKEVPKVRGRGRPPKVKITEL-----LNKTDNRPLKKL 669
Db 415 QKSTGNSKPKGLKTKYKMKQMTLLDMAKGTQ---KMTAPRNSGGTPTSSKPHKL 471
Qy 670 EA-----QETLNEEDKAKIAKKOMRQKQVORGECLTTIOGAARKKQKQETSKLH 720
Db 472 PPAALHLIAYYKKNKREDKRSALSCVISTARTLLSSEDRARLPEELRSLVQKRYELLEH 531
Qy 721 KE--AKKSAEKEGKTKQEKLEKVKREKEKVKKEEVTAKPACADKTLATQR 778
Db 532 KRWASMSBQKRYLKKKREELKKLKEKAKER---REKE----- 569
Qy 779 RLEERQOQMILEEMKPTEDMCLTDHOPLPDFSRVPGLT-LPSGAFSDCLTIVEFLHSF 837
Db 570 MLELEKQ-----KRYEDQELTG-KNLPAFLVDTPEGLPNTLFGDVAMVVEFLSCY 620
Qy 838 GKVLGFPAPKDVPSGLVQEGLLCQSDSLGVEQDILLVRLKAAALHDPFPYCSQSLKILG 897
Db 621 SGLLLPDAQYPTAVS-LMEALSADKGGFLYNLRVLVILLQTLQTLQDLQDAIEDYGBLG 679
Qy 898 EKVSEIPLTRDNVSEILRCFLMAYG-----VXPALCDRLRTQ 934
Db 680 MKLSEIPLTHSVSELVRLCLRRSDVQESBGSDDTDNKKDSAAFPEDNEVQDEFLEKLETS 739
Qy 935 PFOAQPOOKAAVLAFPPVHELNGSTLIINEIDKTLSSMSYRKWKWIVVEGRLLRLK----- 990
Db 740 EFFELTSEEKQILTALCHRI-----LMTYSVDHMETRQMSAELW--KERLAVLKEEND 793
Qy 991 TVLAKRTGRSVEV-----GRPEECJGRRRSRRIEETSGMEEBEESIAAIVGRR--- 1042
Db 794 KKRAEKQKRKEWAKNKENGKVENGLGKTRKRIKVPFQPDVTEAEDMISAVKSRLLA 853
Qy 1043 --GRDGEVDATASSIPELERQIEKLSK-----QLFFRKLLHSSQMLRAVSLGQDQRY 1094
Db 854 IQAKEREIQEREMKV-KLERQAEERIRKHAAAEKAFQSGIAKAKLVMRTPTIGTDNR 912
Qy 1095 RRRYVWL-PYLAGIFVEGTEGNLVPVEVKKETDSLVKVAHAASLNPALPFSMKMELAGSNT 1153
Db 913 HNRWLFSDVPGFLFIE-----KGWVHDSID----- 938
Qy 1154 TASSPARARSPLTKTKCFMQPRHFKS-PVRGQDSEQPQAOQLQPEAQLHVPAPQPOQL 1212

Db 1036 IHL--ARKPNLGLKSCDGNQ--LNLFLRSLDLIEVATRLQKGLGVYETSBEFARVIS-- 1090
Qy 1506 LQIRGWTCPSPDSTREDLAYCEHLSDSQEDITWRGREGGLAPQKTTNPLDLAVWRLAA 1565
Db 1091 -----LEKLKDFGECV-----IA 1103
Qy 1566 LEQNVKRRYLREPLW-----THEVVLEKALLSTPNGAPEGTTTISYBIT 1611
Db 1104 LQASVIKFLQGFAPKQRRKLQSEDSAKTEEVDEEKKMVE-----EAKVA 1150
Qy 1612 PRIIRWOTLQCRSAAHVCLGLHLERSIAWEKSVNVTCLVCRKGDNDREFLLLCDGCD 1671
Db 1151 SALEKWKTAIRBAQTFSRHVLGLMLDACIKWMSAENARCKVCPKGGEDDKLLDCEN 1210
Qy 1672 RGCHYCHRPMEAVPEGDWFCV-----LAQVVEGFEFTQKPGPK 1713
Db 1211 KAFHLFCLRPALYEPDGEWQPCQACQATARRNSGRNVTESASEDSEDESEEE 1270
Qy 1714 RGQKKSYSLNFSGDGRRRLVLKGRES---PAA--GPR-----YSEERLSPSK----- 1759
Db 1271 EEEEEEDYEV---AGLRPRKTIIRGKHSVIPPAAARGRRPGKPKHSTRSQPKAPPVD 1327
Qy 1760 -----RRRLSNRHSHDLTFCEIILMEMESHDAAPFKPEVNPRLVSGYRRIKNP 1810
Db 1328 DAEDVELVQTKRSRRSRLQKCEIILHKIVYRFSFPREPVTDRDEADYDVITHP 1387
Qy 1811 MDFSTMRELLRGGYTSSEFAADALLVFDNCQTFENEDDSEV 1852
Db 1398 MDFQTVQNKSCGYSRSVQEFELTDMKQVTTAEVYVNCRGSHV 1429

RESULT 36
US-09-839-479-65
; Sequence 65, Application US/09839479
; Publication No. US20020039779A1
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042002
; CURRENT APPLICATION NUMBER: US/09/839,479
; CURRENT FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 65
; LENGTH: 176
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-839-479-65

Query Match 4.8%; Score 472; DB 9; Length 176;
Best Local Similarity 54.5%; Pred. No. 4.8e-18;
Matches 96; Conservative 30; Mismatches 50; Indels 0; Gaps 0;
Qy 780 LEORQKQOMILEEMKKPTEDMCLTDHQPDPFSRVPGTLTSPGAFSDCLTIVEFLHSGK 839
Db 1 LEORLELEMAKELKKPNEDMCLADQKPLPRLPGLVLSGSTFSDCLMVVQFLRNF GK 60
Qy 840 VLGFDPKDVPSLGLVQEGLLCQGSGLGEVDLLVRLKAAHDPGFFSYCQSLKILGEK 899
Db 61 VLGFVDNIDVNLVQEGLLNIGDSMGVQDILLVRLLSAAVCDPGLITGYKAKTALGEH 120
Qy 900 VSEIPLTRDYNVEILRCFLMAYGVKPCALCDRLRTQPFQAOQPOOKAAVLAPPVHEL 955
Db 121 LNVGVNRDYNVSEILOIFMEAHCGQTETLSLTKAFOAHTPAQKASVLAFLINEL 176

RESULT 37
US-10-376-537-66
; Sequence 66, Application US/10376537
; Publication No. US20030224405A1
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042001
; CURRENT APPLICATION NUMBER: US/10/376,537
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US/09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 66
; LENGTH: 176
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-376-537-66

Query Match 4.8%; Score 472; DB 15; Length 176;
Best Local Similarity 54.5%; Pred. No. 4.8e-18;
Matches 96; Conservative 30; Mismatches 50; Indels 0; Gaps 0;
Qy 780 LEORQKQOMILEEMKKPTEDMCLTDHQPDPFSRVPGTLTSPGAFSDCLTIVEFLHSGK 839
Db 1 LEORLELEMAKELKKPNEDMCLADQKPLPRLPGLVLSGSTFSDCLMVVQFLRNF GK 60
Qy 840 VLGFDPKDVPSLGLVQEGLLCQGSGLGEVDLLVRLKAAHDPGFFSYCQSLKILGEK 899
Db 61 VLGFVDNIDVNLVQEGLLNIGDSMGVQDILLVRLLSAAVCDPGLITGYKAKTALGEH 120
Qy 900 VSEIPLTRDYNVEILRCFLMAYGVKPCALCDRLRTQPFQAOQPOOKAAVLAPPVHEL 955
Db 121 LNVGVNRDYNVSEILOIFMEAHCGQTETLSLTKAFOAHTPAQKASVLAFLINEL 176

RESULT 38
US-10-702-148-65
; Sequence 65, Application US/10702148
; Publication No. US20040063145A1
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042002
; CURRENT APPLICATION NUMBER: US/10/702,148
; CURRENT FILING DATE: 2003-11-05
; PRIOR APPLICATION NUMBER: PRIOR APPLICATION NUMBER: US/09/839,479
; PRIOR FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 65
; LENGTH: 176
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-702-148-65

Query Match 4.8%; Score 472; DB 15; Length 176;

;
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042002
; CURRENT APPLICATION NUMBER: US/10/702,148
; CURRENT FILING DATE: 2003-11-05
; PRIOR APPLICATION NUMBER: PRIOR APPLICATION NUMBER: US/09/839,479
; PRIOR FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 51
; LENGTH: 59
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-702-148-51

Query Match 3.1%; Score 308; DB 15; Length 59;
Best Local Similarity 98.3%; Pred. No. 9.7e-10;
Matches 58; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1788 AWPEKPNRLVSGYRRIKNPDMDFSTMRELLRGYTSSEFEAADALLVFDNCOTFN 1846
Db 1 AWPFLEPNRLVSGYRRIKNPDMDFSTMRELLRGYTSSEFEAADALLVFDNCOTFN 59

RESULT 44

US-10-415-187-5

; Sequence 5, Application US/10415187

; Publication No. US20040044184A1

; GENERAL INFORMATION:

; APPLICANT: BAUGHN, Mariah R.

; APPLICANT: YAO, Monique G.

; APPLICANT: WALIA, Nandinder K.

; APPLICANT: GIETZEN, Kimberly J.

; APPLICANT: THANGAVELU, Kavitha

; APPLICANT: LU, Yan

; APPLICANT: DING, Li

; APPLICANT: YUE, Henry

; APPLICANT: TANG, Y. Tom

; APPLICANT: LAL, Preeti G.

; APPLICANT: BATRA, Sajeev

; APPLICANT: LU, Dyung Aina M.

; APPLICANT: SANJANWALA, Madhu S.

; APPLICANT: ARVIZU, Chandra

; APPLICANT: RAMKUMAR, Jayalaxmi

; APPLICANT: GRIFFIN, Jennifer A.

; APPLICANT: GURURAJAN, Rajagopal

; APPLICANT: AZIMZAI, Yalda

; APPLICANT: XU, Yuming

; APPLICANT: BURFORD, Neil

; TITLE OF INVENTION: CYTOSKELETON-ASSOCIATED PROTEINS

; FILE REFERENCE: PF-0828 USN

; CURRENT APPLICATION NUMBER: US/10/415,187

; CURRENT FILING DATE: 2003-04-23

; PRIOR APPLICATION NUMBER: PCT/US01/50983

; PRIOR FILING DATE: 2001-10-26

; PRIOR APPLICATION NUMBER: US 60/244,022

; PRIOR FILING DATE: 2000-10-27

; PRIOR APPLICATION NUMBER: US 60/247,370

; PRIOR FILING DATE: 2000-11-08

; PRIOR APPLICATION NUMBER: US 60/251,831

; PRIOR FILING DATE: 2000-12-07

; NUMBER OF SEQ ID NOS: 28

; SOFTWARE: PERL Program

; SEQ ID NO 5

; LENGTH: 2803

; TYPE: PRT

;
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20040044184A1 5844189CD1
US-10-415-187-5

Query Match 3.1%; Score 305; DB 15; Length 2803;
Best Local Similarity 19.2%; Pred. No. 2.1e-07;
Matches 375; Conservative 201; Mismatches 674; Indels 700; Gaps 89;

Qy 137 SSPSHNTNLRAGSOKFWANGTHSPMGLNFDQSLEYLDFPDQNF-----EVCSGI 186
Db 874 SIPSRTTEATQGLDYVPFSAGTISPTS-SLEBDKGFSPPCEDFSVTGSEKRGIGL 932
Qy 187 HPDEAAEKEMTSVVAENGTLVCSLELEEKQPELKMCCYNG----SVPSVESLH- 236
Db 933 SGERAVBEE-----EBET---ANVMSE---KLCQYGTVPFSAAG-HALHPGEPAL 977

Qy 237 ---QEVSVLVPDPTVSCLDOPSHLPDQLEDTP----- 265
Db 978 GEAEERCLSPDDSTVKVASPPPGPPSATHTPFFHQSPVEEKSEPDQFQEAADSGDTKTP 1037

Qy 266 -ILSEDSLEPNSLAPFVGGLYGIDDTLMGAEDKPLXDSPIVIALDCP-----SLN 319
Db 1038 GVGKEDAAE--ETVKPGPEGTL-----EKEEKVPPRSP--QAQEAAPVNIDEGLT 1084

Qy 320 NATAFSLIADDSQTSISIF-ASPTSPVILGESVLQD--NSFDLNGSDAEQEEEMETQSSD 376
Db 1085 GCTIQLLPAQDKATVFEIMEAGEPTGPILGAEALPGGLRTLPQEPGPKQKDEVILRYPDRS 1144

Qy 377 FPPS-----LTOPAPDQSS-----TIQ-LHPATSPAVSPPTTSPAVSLVWSPA 417
Db 1145 LSPEDAESLSVLVSPSPDTANQEPKSPCLTQYLHKORWPEVSEDTQSL- 1199

Qy 418 ASPEISPEVCPAATVSPVAVFVSPASSAVLPAVSLVDP-----LTASVTSPKASP 470
Db 1200 -----SESPSKETSLDVS-SKQSPESLGTQLFGELNLGKEEMGHLMQAENTSHHTAP 1252

Qy 471 VTSAPAAFPPTASPANKDVSSFLETTADVEITGGL-----TASGSDVMRRR 518
Db 1253 MSVPEPHAATASPTTGTTRYSAQT---DITDSDLRKSPASSFSHSTPSGNGKYLPGA 1308

Qy 519 IATPEEVRLPLQHGWRREVRIKKGSHRWQGETWYYPGCKRMKQPFQKYLKSLRNLVHSV 578
Db 1309 ITSPDEHILT-----PD----- 1320

Qy 579 RREHFSFSPMPVGVDFEERDTPPEGLQWVLSASEIISRIOAITGKGRPRNTKAKTYKE 638
Db 1321 --SSFSKSPESLPGPALED---IAIKW---EDKVP-----GLKORTSEQKKEP 1360

Qy 639 VPVKRGGRPPVKITELLNKTNRPLKLEAQTNEEDKAKIAKSKKMRQKVQGE 698
Db 1361 EPKDEVILQKDKTLEHKEVVEPKDTAIYQDEALHVKNVAVKQDQKALEQGRDLQKDT 1420

Qy 699 CLTTIIOGARNRKQETKSLKHKKKSAEKKGTQKOEK---KEKVREKKEKVKM 755
Db 1421 AL-----EQDKAL---EPKDKLEKKKALEQKDKIPEEKDKALEQKDTALEQ 1466

Qy 756 KKEEVTAKPACAKDTLATQRRRL--EERQK---QQMILEEMKKPTEDMCLTDHQLPDF 811
Db 1467 KDKALEPKDKLEQKDRVLEQKEKIPEEKDKALQKVRSEVHKAPEDTVAEMKDRDLQET 1526

Qy 812 SRVPGTLTSGAFSDCLTIVEFLHSGKVLGDFDPAKDVPSLGVLOEGLLCOGDSLGEVOD 871
Db 1527 DKAP-----EQKHQ-----AOE-----QKDKVSEKKD 1548

Qy 872 LLVRELLKAALHDPGFPSPVCSKLTILGKVSSEIPLTRDNVSEILRCFLMAYGVXPALCRL 931
Db 1549 QALE-----QYVWALGQK-----DEA 1564

Qy 932 RTQPFQAPPPQKAAVLAFPPVHELNGSTLIINEIDKLTLESMSYRKKNWIVEGRRLUKT 991
Db 1565 LEQNIQALEEN-----HOTQEQESLVQE--DKTRKPKMLBEKSPKVKAMEEKLEA 1613


```
QY 992 VLAKRTG-----RSEVMGRPRECIGR---RRSRIME---ETSGMBEEEEEEIAAVPGRRG 1043
Db 1614 LLETKALGLBESLVQEGRAREQEKEYRGQDVQVQEWQETSPTREEPAGEQKELAP-----1669
QY 1044 RRDGEVDATASSIPELEGRQIEKLSKROLFFPKLLHSSQMLRAVSLGQDRY-----RRR 1097
Db 1670 -----AWEDTSPEQDNK-----YWGR-----EDVALEQDITVWRELSCEK 1705
QY 1098 YWVLPYLAG-----IFVBGTGNLVPEEVIKKTDSLKAAHASLNFALFMSKMLAGSN 1152
Db 1706 WVFPHELDQCARPHYTERESTFLDEG--PDDQEVPLREHATRSWASDFK-----D 1757
QY 1153 TTASAPAR--SRPLTKYGFMPORH-----FKSPVR-----GQD 1186
Db 1758 FQESSPQKGLEVRWLAESPVLPEEDKLTRSPFEIISPPASPPEMVGQRPSPACQE 1817
QY 1187 SEQQAOLQPEAQLH-----VPAQPOPOLQLOLQH-- 1217
Db 1818 SPIDPKLPMHMKNEPTTPSWLADIPWVPWKDRPLPPAPLSPAGPPTP-----APESHTP 1873
QY 1218 -----KGFLEQEGSPLS--LGOSQHDLS---QSAFLSWLSQTQSHSSL 1255
Db 1874 APFSWGTAEYDSVVAQVEGAEEGGYPYPLGKDYRKAEGEREEGRAEAPDKSSHSSK 1933
QY 1256 LSSSVLTPDSSFGKLDPAPOPEPEPEDEAESPDLOAFWF--NISAQM-----1303
Db 1934 V-----PEASKSHATTEPEQ--TEPEQRETPYPDERSFOYADIEQMMLTGLGPACPT 1985
QY 1304 -----PC-----NAAPTPLAYSEDOPT-----P 1322
Db 1986 REPPLGAAGWPPCLSTKKEAAGRNTSAEKELSPISPKSLOSDTPTFSYAALAGTVPVP 2045
QY 1323 SPQQLASSKPNRPSAANPCSPV--QFSSTPLAG---LAPKRAGDPGEMPQS-----1370
Db 2046 RPEGCPSEPSLTTPAVPPRAPILSKGPSPLNGNILSCSPDRSPSPKESGRSHWDST 2105
QY 1371 -----PTGLQPKRRGRPPSKFKQMEQRYLTQTAQVPPEMCSGWWITPDPEMLDAM 1424
Db 2106 SDSELEKGAREQKEQAQSPS-----PPHPIMPGSPTLM--PETE-----2143
QY 1425 LKALHPGIRKALKHLKLRDPLQEVCLRPSADPIPEPQLPAFQEGIMNSPKKEYTY 1484
Db 2144 ---AHV-----SPPLDLSHLP-----ARPSLD-----FPASAFGSSLQAPP-- 2178
QY 1485 ETDLAVLQWVELEQRVIMSDLIQIRGWTCPSPDSTREDLAYCEHLSDSOEDITWRGPGRE 1544
Db 2179 -----OLPFAPEPRS--APCGSLAFSGDRALALAPG-- 2207
QY 1545 GLAPQKRTNPLDLAVMRLAALAEQNVKRRYLREPLWPTHEVYLEKALSTP---NGAPEG 1601
Db 2208 ---PPTRTRHDEYLEVTKAPSLDSSLQ--LPSPSSPCAPLLSNLPRPASPALSEGSSSE 2262
QY 1602 TTTEISYBITRIRIWRQTLOR-----CRSAHVCLCL-----GHLERSIAWEK 1645
Db 2263 ATTVISSVAERFSPSLAAEQSEGELDPGMEPAHSLWDLTPLSPAPPASLDLALAPAP 2322
QY 1646 SVNKVTCVLCRGDNDPEFLLLCDGCDRCCHYVC-----HRPKWEAVPBGDFWCTVCLAAQ 1700
Db 2323 SL-----PGDMGDGIL-----PCHLECSAATEKPSPTQVPSED-----CAANG 2361
QY 1701 VEGEFTQKPG--FPKRQKRRKSGYSLNFSEG---DQRRRR-----VLLKGRE---SPAAG---1747
Db 2362 PTETSPNPPGPAKAENEAAACPAWERGAWPEGAERSSRPDTLLSPQVPCFAGGSGG 2421
QY 1748 -----PRYSERLSPS 1758
Db 2422 PPSSASPEVAGPOGCATEPRPHRGELSPS 2451
```

RESULT 45

US-10-437-963-144559

; Sequence 144559, Application US/10437963

```
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 144559
; LENGTH: 1855
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_45361C.1.pap
; US-10-437-963-144559
```

Query Match 3.08; Score 295.5; DB 16; Length 1855;

Best Local Similarity 19.64; Pred. No. 4.1e-07;

Matches 375; Conservative 226; Mismatches 645; Indels 667; Gaps 89;

```
QY 187 HPDSEAEKEMTSVVAENGTLG---VCSLEEXQPELKMCGYNGSVSPVSELSHOEVSVLV 243
Db 80 YPNETMRAELSVKLGLTDRQLQMWFCRRLLKDRKPP-----PKRQOLEEVPVPV 129
QY 244 PDPTVSCLDPPSHLPDQLEDTPLSEDSLEFPFNSLAPEPVGSLGYIDDTLMLGAEDKLP 303
Db 130 MAP-----PPVLPPLP-----PHSELTMG--TGMVG-----154
QY 304 LXDSPTVLSALDCPSLNATAFLLADDSQTSSTSFASPTSPVPGESVLQDNSFDLNNGS 363
Db 155 ---EPLLS---PSSRRGT-----GRPSAVPRISASEVARRYE-----186
QY 364 DAQEEMETQSDPPPSLTOP--APQDSTTQLHPATSPAVSPT-----TSPAVS 411
Db 187 -----PPQVMLPPLAP-----MQLTRAHRVIDSVLERLIGEPLREDGPIVG 227
QY 412 LVWSP-----AASPEISPEVCPAASTVVSVAVFS---VVSPPASSAVLPAVSLVPLTASVT 464
Db 228 VEPDPLPGAFAPIVPEQQRQPPRSYEAKMFGSDTKHIKASAFLSID---FPVPSTV 284
QY 465 SPKASPVTSAAAPPTASPANKOVSSFLETTADVEETGEGLTAGSGDVMRR-----RIA 520
Db 285 SGKSKMTGSSSHL--GSRVREYQFIPEOSSDIYERT---TQSRYYDVLTEASDSRMS 338
QY 521 TPEEVRPLQHGWRREVRKKGSHRMOGETWYGPCGKRMKQFPPEVIKLSRNLVHSVRR 580
Db 339 TGS-----RFLHGVEKAPGYTFHGQISGS-----362
QY 581 EHFSSPRMPYGDPFPERDTPGELQWLQSAEEIPSR-----IQAITKRG--RPRNTEKA 634
Db 363 SHLSLGRPPV---FSSGSTDYEMDPSNINSSSVNESQYGIPOVAGPENSVAPSDRMVY 419
QY 635 KTEKVPVKVGRGRPPPKVKITELLNKTDNRPLKLEAQETNEEDKAKIAKSKKMRKV 694
Db 420 HDDDAFRMERKRKHSEEVKIAKEVEAHEKRIRKELEKQDMLNR-----KRESEVREM 472
QY 695 ORGECLTTIQOARNEKQETKSLKHKEAKKSKXAEKEGKTOKKLEKVKREK---KE 751
Db 473 ERND-----RERKKEERLLUREK-----QKEERFQREKREHMEKYLKQ 515
QY 752 KV---KMKKEEVTKAKPAC--KADKTLATORR-----LEBRQOQMILEMKKPKTE 798
Db 516 SLRAEKMRQKEELEKEAARQAANERATARRAREYMELMEDERLELMELVSRSKGLP 575
QY 799 DMCLTDHQLPDPDFSRVPGL-----TLPSGAFSDCLTIVEFLHSFG 838
```

```
Db 576 SMLSLSDTLQQLDSFRGMLRQFSEIVRLKVPFSIKPWTSSDNIGNLLMWWFFITFA 635
Qy 839 KVLGDPKAVPSLGVLOEGL--LCOGDS--LGEVQDLLVRLKLAALHD-----883
Db 636 DVLG-----IPSF--TLDFVQSLHDVDSRLGE--LHVALLKSIINDIEDVARTPSVA 685
Qy 884-----PGFP-----SYCOSLKITLGEKVEIPLTRDNVSEILRCFLMAGVAPALCDR 930
Db 686 SCMTANPGGHPQIVEGYDNGFNILAWQRHNLTT---WPEILRQFGLSAGLQPOLRXR 742
Qy 931 LRTOFQOAPQQAVALAPFVHELNGSTLIINEIDKLTLESMSYRKNKMWIVEGRRLRLK 990
Db 743-----NAENVNHDN-----EGR-NGED 760
Qy 991 TVLAKRTGRSVMG--RPESCLGRRSSR-----TMEETSGMEEREE 1032
Db 761 VISILRSGSAVNAAKMKRGYGNRRSRHRLTPGTVKFAAFHVLSEGSQGLTILEVA 820
Qy 1033 ESIAAVPGRRRRDEVDATASSIPELERQIEKLSKROLF-----FRKLLH 1079
Db 821 EKI-----QKGLRDLTTKTPPEASISAALSRSKLFERTAPSTVCVTKPYRKOPAD 872
Qy 1080 SSOMLRAVSLGQDYYRRYVWLPYLIGIF-----VFQNTISECEVEKDVDAERDESECDADDD 1108
Db 873 SEAVLAA--AREKIR-----VFQNTISECEVEKDVDAERDESECDADDD 918
Qy 1109 VEGTEGNLVPEV-----IKKETDLSKVAHAASINPALFMSKMELAG 1150
Db 919 PDGDEVNIEKDVKTSLVKAQDGMPTAVGDIKKETNSI-----VNSLTTPLIH 967
Qy 1151 SNTTASSPARSRPLTKPFQMPRHPKSPVRQDSEQPOAQLOPEAQLHVPAQPOQL 1210
Db 968 TKSESSESLRDLKSVQVRTTSDLPAEISS-----DNHEGASDAQDAIDESNQGESWV 1022
Qy 1211 QLQLQSHKGFEOBGSPLSLQSOHDLSSQAFSLSWLSQTSQSHSLSSVLTPOSSPGKL 1270
Db 1023 Q-----GLAEGDYCDLSVERLNAL--VALIGVATEGNISIRAVLEE-----RL 1063
Qy 1271 DPAPS--OPPPEEPDEAESDLOQAFWFNISQMPCNAAAPTPLAVSEDPQTPPSPOQL 1327
Db 1064 EAASALKQMAEAQLDKRRSREBFS--KMQYDSGMGLKTDVDQNTLAEISNLTPVHNLV 1122
Qy 1328 ASSK-----PMNRPSAANPCSPV-----QFSSTP--LAG-----LAKPRA-- 1361
Db 1123 KDSNGSLVNNELFVQDQSPNACSVVHERNGVRQBFSANPENLSGQQYVTSKRSQL 1182
Qy 1362---GDPE-----MPQSPTGLQPKRRGRPPSKFFKQMEQRYLTQLTAAQVPP--- 1406
Db 1183 KSYIGHKABQLHVYRSLP-----LGQDERRN-----YMQPSTASPDGSGR 1226
Qy 1407---EMCGSWWIPDP--EMLDAMLKALHPRGTREKALKHLNKHKRFLOE-----VCLR 1455
Db 1227 IFPESRDGYWRLIDSJETFDALVSLDTRGIRESHLSMLQSIPTFEAIGRKRCASIE 1286
Qy 1456 PSA-----DPIFEERQLPAFOEGIMSWG-----PKEKTYETDLAVLOWVELEQVFI- 1502
Db 1287 PSAGRVLKNGTSEIISPNHSEFSPCSTLSGVTADSAMAYSDFSRIELGRNDVEKTAIS 1346
Qy 1503--MSDLQIRGW-----TCPSPDSTR--EDLAYC-----1526
Db 1347 ERADLFIK-WWMEKCNHQPCTAMKHGKRCSELIQCDDFCYQIYLAETHCASCHKTFK 1405
Qy 1527-----BHLSDSQE-----DIWRGFGREGLAPOQRKTNPLDLAVMRLAALBQNVKRYLR 1576
Db 1406 SIHNISEHSSQCEBKRRTPDNKMQISDYSVPV-----GLALLKLLATVEASVPAEAL- 1459
Qy 1577 EPLW-----PTHEV-----VLEXALLSTPNGAEGTTEISYITPRIR 1615
Db 1460 EPFTVDVYRKGWVKLYSTSTKVFEMLTITLEGAIIRDRLFUSSDFETTEL-LNLSTQDS 1518
Qy 1616 IWRTQLOCRSA-----AHVCLCLGHLERSIAWEKSVNVTCLVCRKGDNDDBLL 1665
```

```
Db 1519 ASRNTVPRSGADVLPWVPDPTVAAVLRLDLLDSAISY-----TLRKVGSNKE--- 1567
Qy 1666 LCDCCDRGCHYCHRP-----KMEAVPEGDFCTVCLAQOQVEGFTQKPGFPKQX 1717
Db 1568-----RGAGFEMKLPPTYTPAKTKQETPMGTGF-----DROETWLTPSNGRGRGR 1615
Qy 1718 RKSGLYNFSEGDRRRRLVKGRESF---AAGPR-----YSEERLSPSKRR 1762
Db 1616 GRG-----SRGSRGRSRSGKVPGRGISSSPKIEFRGYSASAVSEKAPR 1662

RESULT 46
US-10-755-889-615
; Sequence 615 Application US/10755889
; Publication No. US20040171823A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES ASSOCIATED WITH THE NF-kB
; FILE OF INVENTION: PATHWAY
; FILE REFERENCE: D0284 NP
; CURRENT APPLICATION NUMBER: US/10/755,889
; CURRENT FILING DATE: 2004-01-13
; PRIOR APPLICATION NUMBER: U.S. 60/440,068
; PRIOR FILING DATE: 2003-01-14
; PRIOR APPLICATION NUMBER: U.S. 60/469,757
; PRIOR FILING DATE: 2003-05-12
; NUMBER OF SEQ ID NOS: 823
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 615
; LENGTH: 2468
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-755-889-615
```

Query Match 2.9%; Score 288; DB 16; Length 2468;

Best Local Similarity 19.2%; Pred. No. 1.5e-06;

Matches 326; Conservative 249; Mismatches 583; Indels 544; Gaps 82;

```
Qy 164 NFDQELYDSFPDQNFVEVCVGIHPDEAAEKEMTSVVAENGTLVCSLELEEXQPELMKC 223
Db 156 SFSQNFIEIFTDOEIGELLSTTHP--ANKASLTFCEEGDWKNSLDRHNLQDFINIK 213
Qy 224 GYNGSV-PSVESLHQEVSVLPDPTVSCLDPSHLPOLEDTPILSDSLE---PFNSLA 279
Db 214 LNSASILPEMEGLS-----EFTEYLSE-SVEVPSFFDILE 247
Qy 280 PEPVSGG-----LYGIDDTLM--GAEDK-----LPLXDS 307
Db 248 P-PTSGGFLKLSKPCCVIFPGRGDSALFAVNGFNMLINGGSRKSCFWKLIRHLDRVDS 306
Qy 308 PVTISAL---DCPSLNATAFSLADDSTSTSPASTSPVLGESVLQDNSFDLNNGSD 364
Db 307 ILLTHIGDDLNLPGIN-----SMLQRKIAEL-----E 332
Qy 365 AEOBEMETQSSDF-----PPSLTOPAP-----DOSSTIQ---LH 395
Db 333 BEQSQGSTTSDWKNKLSIPDLGVFLNVLENLKNPEFNKMKRSIEBEAFTLYLNKLS 392
Qy 396 PATSP---AVSPTTSPAV-----SLWVSPA-ASPEISPEVCPAASTVSPAVFS 440
Db 393 MKPEPLFRSVGNTIDPVLFOKMGVGKLEVMYVLNPKVSKSEMQYFMQQTGNTKDKAEF- 451
Qy 441 VVSPASSAVLP---AVSLEVLPTASVTSKASPVTSAPAAFPPTASPAKDVSSFLETTAD 497
Db 452-----ILFNGQEVDPISYLTSSV-----LIVMHPANPAEKIRVLPFGNST 494
Qy 498 VEITIGLGTASGSDVMRRRIATPEBVRPLQOHGWRREVRIKKGSHRWQOGETWYVPCG 557
Db 495 QYNIL-EGLEKLKHLDFLKQPLATQKDLTGQVTPPVVKQTKLKORADRES-----LKPAA 549
Qy 558 KRMKQPFVVIKYLRSNLVSHVSRREHFSPRMPVGDFFEBRDTPTEGLQWVLSABEIPSR 617
Db 550 KPLPS-----KSVRKE-----SKEETPE---VTKNVHVEKPPK 579
```


Db 453 TKPLSSK---SVRKE-----SKEAPEATKA-----S 476
Qy 627 RPRNTEAKTKE---VPKVKGR-GRPKVKITELLNKTNRPLKLEAQFTLNEEDKAK 682
Db 477 QVEKTPKVESKEKVIIVKDKPGKVESPSVTEKVPKESQSPV-KAEVAEKAATESKPK 535
Qy 683 IAKSKKQNRKORVORGECLTTTQGOARNRKRQETSLKHKEAKKSAEKEKGT-----736
Db 536 VTOKKV-----KKEIKTPKEKKEKPKKEVAKEDKTPLKDE 575
Qy 737 --KOEKUKKVKRKEKVKVKEEVTAKAPACKADKTLATQRRLEERQKQOMILEMK 794
Db 576 KPKEEAKKEIKKEIKKEKELKVKETPLDKAKEVKKDEKKEVKEKKEPKKEIK 635
Qy 795 KPTEDMCLTDHQPLDPFSRVGLTLPSCAFSDCLTIVEFLHSFGKVLGPDPAKDVPSLG 854
Db 636 KISKD--IKKSTPLSD-TKPPAALKPKVAKKE-----EPTKKEP-----671
Qy 855 LOEGLLCQDGLGEVQDLLVRLKLAALHDPGFPSCQSLKILGE--KVSEIPLTRDNVSE 912
Db 672 -----IAAGLKDXG-----KVVIKKEGKTEAAATAVGTAA 704
Qy 913 ILRCFLMAYGVX---PALCDBLRTPQFOAQPPOQAAVAPVHELINGSTL-IINEIDKT 968
Db 705 V-----AAAAGVAASGPA--KELEAERSLMSSPED-----LTKDFEELKABEIDVAKDIKPQ 754
Qy 969 LESMSYRKNIWIEGRRLRLKTLVAKRT-----GRSEVENG-----R 1006
Db 755 LELIEDEBKLETBPGE-----AYIQKETEVSKGSAESPDRGITTTEGECEQTPBELE 810
Qy 1007 PEECLGRRRRSRIWETSMEEEEE---EBSIAAVPGRRGRDRGDVATASSIPELERQ 1062
Db 811 PVEKQGVDDIEKFDEGAGFSESSEAGDYEKATEEAEEPEEDGEDNVSGASAKHSPTE 870
Qy 1063 IEKLSKRQLFFRKLLHSSQMLRAVSLGQDRYRRRYWVLYLAGIFVEGTGNL-----1116
Db 871 DEETAKAED-----VHIKEKRESVASGDDRAE-----EDMDEALEKGEAE 911
Qy 1117 VPREEVIKKTDSLKVAAHASLNP-----ALFSKMELAGSNNT-----ASSPARARS 1163
Db 912 QSEEEGEEEDKAEDAREEDHEPKTEABDYVMAVVKAAAGVTEQDYDFLGTAPK---968
Qy 1164 RPLKTKPGFMOPR-----HFKSPVRGQDSE---QPQALQPE-----AQLH 1201
Db 969 -----QFGVQSPSREPASSIHDETLPGGSESEATASDEENREDQPEFTATSGTQTIE 1023
Qy 1202 VPAQFPQLOLQSHKGFLQEQSP--LSLGQSHDLQSQAFLSWLSQTOSSHSLSS 1259
Db 1024 ISSEPTP-----MDMGTPRDVMTDETNNBETESPSQBFVNITKYESSLYSQE 1071
Qy 1260 VLTP-----DSSFGK-----LDPAPSQPPEE-----PEPDE 1285
Db 1072 YSKPVVASFNGLSGSKTDAITDGRDYNASATISP-PSSMEEDFKSALRDAYRPEETD 1130
Qy 1286 AESSPDLOAQFENISQAQPCNA---APTPLAVSEDOPTP-----SPQOLASSKP 1332
Db 1131 VKTGAELDIKDVSDERLSPAKSPSLSPSP---SPIETPLGERSVNFSLTPNIEIKASAE 1187
Qy 1333 MNRPSAANP-----CSPVQFSGSTPLAGLAPKRRAGDGPMPQSP 1371
Db 1188 GEATAVVSFGVTVQAVVEHCASPEEKTLEVVSQSP-SVTGSAGHTPPYVQSPPTDEKSHLP 1246
Qy 1372 TGLQPKRRGRPPKFFKQMKQRVLTQTAQVPPEMCSGWWIIPDPBMLDAMIKAHPR 1431
Db 1247 TEVTENAQAVPSFEFTAKDNERSSISPMDPE-----VPDSF--SPIEKVLSLP- 1294
Qy 1432 GIREKALHKLHKLKRDPL---QEVCLRPSADPI-----FEPQLPA-----FQE- 1472
Db 1295 -LRSPPLIGSEAYEDFLSADKDALGRKSESPFFKGKKGQGFSDKESPVSDLTSDLQDK 1353
Qy 1473 -----GIM-----SWSPKETTYETDLAVLQVVELEQRVIMSDIQIRGWTCPSPDSTRED 1522
Db 1354 QEEKRAGFIPITKEDFSPKASDAEIMSSQSALALDERKLGDD-----GSP--TQVD 1403

Qy 1523 LA-YCEHLSDSOEDITWRGPGREGIAPQRKTTPDLAVMLRLAALAEQNVKRRYLREPLWP 1581
Db 1404 VSQFGSPKEDTKMSIS-----EGTSDKSAT-PVD-----EGAE 1437
Qy 1582 THEVVLKALLSTPNGA-----PEGTTTEIS 1607
Db 1438 TYSHMEGVASVSTASVATSSFPPEPTDDVS 1467

RESULT 48

US-09-839-479-61
; Sequence 61, Application US/09839479
; Publication No. US20020039779A1
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042002
; CURRENT APPLICATION NUMBER: US/09/839,479
; CURRENT FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 61
; LENGTH: 45
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-839-479-61

Query Match 2.8%; Score 273; DB 9; Length 45;
Best Local Similarity 97.8%; Pred. No. 5.5e-08;
Matches 44; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1652 CLVCRKGDNDPFLLLCDGCRGCHYCHRPKMEAVPEGDWFCCTVC 1696
Db 1 CLVCRKGDNDPFLLLCDGCRGCHYCHRPKMEAVPEGDWFCCTVC 45

RESULT 49

US-10-376-537-62
; Sequence 62, Application US/10376537
; Publication No. US20030224405A1
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042001
; CURRENT APPLICATION NUMBER: US/10/376,537
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US/09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 62
; LENGTH: 45
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-376-537-62

Query Match 2.8%; Score 273; DB 15; Length 45;
Best Local Similarity 97.8%; Pred. No. 5.5e-08;

Matches 44; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1652 CLVCRKGNDRFLLLCDCGRGCHYCHRPKMEAVPEGDWFCCTVC 1696
|||||
Db 1 CLVCRKGNDRFLLLCDCGRGCHYCHRPKMEAVPEGDWFCCTVC 45

RESULT 50

US-10-702-148-61
; Sequence 61, Application US/10702148
; Publication No. US20040063145A1
; GENERAL INFORMATION:
; APPLICANT: Jones, Michael H.
; TITLE OF INVENTION: TRANSCRIPTIONAL REGULATOR
; FILE REFERENCE: 06501-042002
; CURRENT APPLICATION NUMBER: US/10/702,148
; CURRENT FILING DATE: 2003-11-05
; PRIOR APPLICATION NUMBER: PRIOR APPLICATION NUMBER: US/09/839,479
; PRIOR FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 09/418,710
; PRIOR FILING DATE: 1999-10-15
; PRIOR APPLICATION NUMBER: PCT/JP98/01783
; PRIOR FILING DATE: 1998-04-17
; PRIOR APPLICATION NUMBER: JP 9/310027
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: JP 9/116570
; PRIOR FILING DATE: 1997-04-18
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 61
; LENGTH: 45
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-702-148-61

Query Match 2.8%; Score 273; DB 15; Length 45;
Best Local Similarity 97.8%; Pred. No. 5.5e-08;

Matches 44; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1652 CLVCRKGNDRFLLLCDCGRGCHYCHRPKMEAVPEGDWFCCTVC 1696
|||||
Db 1 CLVCRKGNDRFLLLCDCGRGCHYCHRPKMEAVPEGDWFCCTVC 45

Search completed: March 9, 2005, 15:08:43
Job time : 178.703 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 9, 2005, 13:06:37 ; Search time 136.112 Seconds
(without alignments)
5336.298 Million cell updates/sec

Title: US-10-702-148-13
Perfect score: 9826
Sequence: 1 MEMEANDHFNFTGLPPA.....MRRFFSRWEEFYQKQANL 1878

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 65 summaries

Database : A_Geneseq_16Dec04.*
1: geneseqp1980s.*
2: geneseqp1990s.*
3: geneseqp2000s.*
4: geneseqp2001s.*
5: geneseqp2002s.*
6: geneseqp2003as.*
7: geneseqp2003bs.*
8: geneseqp2004as.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	9816	99.9	1878	2	AAW81170 Human BAZ
2	9647	98.2	1878	4	AAW40239 Human pol
3	9590.5	97.6	1873	7	ABM85417 Human pro
4	8666.5	88.2	1727	4	ABM95554 Human pro
5	7897	80.4	1589	4	AAW42025 Human pol
6	6777.5	69.0	1586	7	ABM85416 Mouse pro
7	3777	38.4	779	4	ABM95562 Human pro
8	2766	28.1	524	8	ADP09529 Human pro
9	2761	28.1	555	3	AAW21050 Human nuc
10	2483.5	25.3	1972	2	AAW81171 Human BAZ
11	2482.5	25.3	1972	6	ABR64241 Angiogene
12	2482.5	25.3	1972	8	ADP54420 Human pro
13	1420	14.5	280	6	ABU70931 Human adi
14	1224	12.5	708	4	AAW93073 Human pro
15	1108.5	11.3	796	4	AAW93402 Human pol
16	1108.5	11.3	796	8	ADL30971 Human pro
17	671.5	6.8	1540	2	AAV07734 Human hTL
18	671.5	6.8	1674	2	AAW81169 Human BAZ
19	671.5	6.8	1674	7	ADP69139 Human MP5
20	671.5	6.8	1674	8	ADP12578 Protein e
21	575	5.9	141	3	AAW54103 Human pan
22	561	5.7	244	4	AAW63263 Human bre
23	528.5	5.4	1483	8	ADQ20910 Human sof
24	527.5	5.4	1476	4	ABM58706 Drosophil
25	525.5	5.3	1483	7	ADF69140 Human MP5

26	521.5	5.3	1527	2	AAW81172	AAW81172	Human BAZ
27	512.5	5.2	1531	2	AAW81173	AAW81173	Human BAZ
28	444	4.5	108	4	AAO3463	AAO3463	Human pol
29	413.5	4.2	2907	3	AAW57452	AAW57452	Human tra
30	406.5	4.1	2781	3	AAW57453	AAW57453	Human tra
31	406.5	4.1	2781	8	ADF42724	ADF42724	Human BPT
32	406.5	4.1	2781	8	ADO00984	ADO00984	Human hom
33	406.5	4.1	2781	8	ADQ18653	ADQ18653	Human hof
34	388	3.9	2759	6	AAO16418	AAO16418	Human nuc
35	379	3.9	475	4	AAW93746	AAW93746	Human pro
36	351	3.6	160	4	AAU16212	AAU16212	Human nov
37	351	3.6	160	6	ABU55281	ABU55281	Human nov
38	338	3.4	1050	4	AAW93902	AAW93902	Human pol
39	338	3.4	1050	7	ADB80954	ADB80954	RING-SH c
40	338	3.4	1050	8	ADL32012	ADL32012	Human pro
41	335	3.4	1127	7	ADB80955	ADB80955	RING-SH c
42	316	3.2	759	4	AAW94613	AAW94613	Human pro
43	305	3.1	2803	5	ABB08161	ABB08161	Human cye
44	304	3.1	2274	4	ABB58657	ABB58657	Drosophil
45	300	3.1	2897	4	ABB58514	ABB58514	Drosophil
46	296	3.0	601	7	ADL22668	ADL22668	Human dis
47	290.5	3.0	2768	4	ABB68397	ABB68397	Drosophil
48	288	2.9	2468	6	ABR64281	ABR64281	Angiogene
49	288	2.9	2468	7	ADP62723	ADP62723	Human pro
50	288	2.9	2468	7	ADP62719	ADP62719	Human pro
51	288	2.9	2468	7	ADP62727	ADP62727	Human pro
52	288	2.9	2468	7	ADP62715	ADP62715	Human pro
53	288	2.9	2468	8	ADL12997	ADL12997	Human ste
54	288	2.9	2468	8	ADN05260	ADN05260	Antipsori
55	288	2.9	2468	8	ADR14614	ADR14614	Human NF-
56	288	2.9	2519	4	ABG16636	ABG16636	Novel hum
57	288	2.9	2527	8	ADN04561	ADN04561	Antipsori
58	287	2.9	2622	4	ABG06418	ABG06418	Novel hum
59	285.5	2.9	513	4	AAW00760	AAW00760	Human bon
60	285.5	2.9	2618	4	ABG02135	ABG02135	Novel hum
61	280.5	2.9	645	8	ADP42726	ADP42726	Mouse BPT
62	280.5	2.9	2364	6	ABM04816	ABM04816	Rat micro
63	280.5	2.9	2459	7	ADP62725	ADP62725	Rat Prote
64	280.5	2.9	2459	7	ADP62713	ADP62713	Rat Prote
65	280.5	2.9	2459	7	ADP62717	ADP62717	Rat Prote

ALIGNMENTS

RESULT 1	
AAW81170	
ID	AAW81170 standard; protein; 1878 AA.
XX	
AC	AAW81170;
XX	
DT	05-MAY-1999 (first entry)
XX	
DE	Human BAZ2-alpha protein.
XX	
KW	Transcriptional regulator; BAZ1-alpha; bromodomain; BAZ2;
KW	atypical zinc finger; testis; human; tumour; BAZ1-beta; BAZ2-alpha; drug;
XX	BAZ2-beta; treatment; cancer; proliferative disorder; screening.
OS	Homo sapiens.
XX	
PN	WO9847920-A1.
XX	
PD	29-OCT-1998.
XX	
EF	17-APR-1998; 98WO-JP001783.
XX	
PR	18-APR-1997; 97JP-00116570.
PR	24-OCT-1997; 97JP-00310027.
XX	
PA	(CHUS) CHUGAI RES INST MOLECULAR MEDICINE INC.
XX	
PI	Jones MH;

Qy	1801	SGYRIIKNPMDFTMTERLLRGYTSSEEPADALLVFNCFNEDDDSEVGKAGHMR	1860
Db	1801	SGYRIIKNPMDFTMTERLLRGYTSSEEPADALLVFNCFNEDDDSEVGKAGHMR	1860
Qy	1861	RFESRWEFFYQKQANL	1878
Db	1861	RFESRWEFFYQKQANL	1878

2. **U.S. DEPT. OF JUSTICE**

RESOL 2
AAM40239
ID AAM40239 standard; protein: 1878 AA.

AC AAM40239;

DT 22-OCT-2001 (first entry)

Human polypeptide SEO ID NO 3384-

Human; neutropic; immunosuppressant; cytostatic; gene therapy; cancer;
peripheral nervous system; neuropathy; central nervous system; CNS;
Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
leukaemia.

Homo sapiens.

XX
PN
WO200153312-A1.

PD: 26-JUL-2001.

26-DEC-2000: 2000WO-US034263-

PR 23-DEC-1999; 99U

PR 25-APR-2000; 2000U

PR 19-JUL-2000; 2000U

PR 14-SEP-2000; 2000U

PR 29-NOV-2000; 2000U

PA (HYSE-) HYSEQ INC.

PI Tang YT, Liu C,

PI Zhou P, Goodrich

DR WPI; 2001-442253/47.

XX XX

PT as central nervous system injuries.

PS Example 5; SEQ ID

CC The invention rela

CC immunosuppressant

CC of the invention m

localised neuropathies and central nervous system diseases, such as

lateral sclerosis, and Shy-Drager S

[illegible]

CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders. Note: The sequence data for this patent did not form
CC part of the printed specification

Db 961 IINEIDKLTLESMSYKRNKWTVEGRRLRLKTVLAKRTGRSEVEMEGPEECILGRRRSSRIM 1020
Qy BETSMEEEEESTAAVPGRRGRDGDVDTASSIPELERQIEKLSKROLFFFKLLHS 1080
Db BETSMEEEEESTAAVPGRRGRDGDVDTASSIPELERQIEKLSKROLFFFKLLHS 1080
Qy SQMLRAVSLGQDRYRRRWVLPYLPIYLAGIFVECTEGNLVPEEVIKKTETSLKVAHAASLNPA 1140
Db SQMLRAVSLGQDRYRRRWVLPYLPIYLAGIFVECTEGNLVPEEVIKKTETSLKVAHAASLNPA 1140
Qy LFSMKMELAGNTTASSPARARSPLKTKPGFMOPRHPKSPVRGDSBPQAOLQBPRAQL 1200
Db LFSMKMELAGNTTASSPARARSPLKTKPGFMOPRHPKSPVRGDSBPQAOLQBPRAQL 1200
Qy HVPAQPQOLQLOQSHKGFLEQSGPLSLGSOHDLSSAFSLWSLQTSQSHSLSSSV 1260
Db HAPAQOPQLOLOQSHKGFLEQSGPLSLGSOHDLSSAFSLWSLQTSQSHSLSSSV 1260
Qy LTPDSSSGKLDPAFSPQPEPEPEDEASSPDQAFWFNISAQMPNCNAAPTPLAVSEDP 1320
Db LTPDSSSGKLDPAFSPQPEPEPEDEASSPDQAFWFNISAQMPNCNAAPTPLAVSEDP 1320
Qy TSPSQQLASSKPMRPSAANPCSPVQFSSTPLAGLAPKRRAGDCEWMPQSPGTGLGQPKRR 1380
Db TSPSQQLASSKPMRPSAANPCSPVQFSSTPLAGLAPKRRAGDCEWMPQSPGTGLGQPKRR 1380
Qy GRPSKFPKOMEQRYLTQLTAAQVPPPEMCSGMMWIPPEMLDAMLKALHPGIREKALHK 1440
Db GRPSKFPKOMEQRYLTQLTAAQVPPPEMCSGMMWIPPEMLDAMLKALHPGIREKALHK 1440
Qy HLNKHRDFLOVCLRPSPADPFIPEPRQLPAFQEGIMSPSPKKTETDTLAVLQWVEELEQR 1500
Db HLNKHRDFLOVCLRPSPADPFIPEPRQLPAFQEGIMSPSPKKTETDTLAVLQWVEELEQR 1500
Qy VIMSDLQIRGWTCSPSTREDLAYCEHLSDSQBDITWRGPGRGGLAPQKRTTNPLDLAV 1560
Db VIMSDLQIRGWTCSPSTREDLAYCEHLSDSQBDITWRGPGRGGLAPQKRTTNPLDLAV 1560
Qy MRLAALQNVKRRYLREPLWTHVLEKALLSTPNGAPEGTTTETISYEITPRIRIWRQT 1620
Db MRLAALQNVKRRYLREPLWTHVLEKALLSTPNGAPEGTTTETISYEITPRIRIWRQT 1620
Qy LQRCRAAHVCLCLGHILERSIAWEKSVNKKVCLVCRKGDNDDEFLLCDGCDRGCHYICHR 1680
Db LERCESAQVCLCLGOLERSIAWEKSVNKKVCLVCRKGDNDDEFLLCDGCDRGCHYICHR 1680
Qy PKMEAVPEGDWFTVCLAQVVEGFTQKPGFPKRGKSGYSLNFSGDRRRRLVLLKG 1740
Db PKMEAVPEGDWFTVCLAQVVEGFTQKPGFPKRGKSGYSLNFSGDRRRRLVLLRG 1740
Qy RESPAAGPRYSEERLSPSKRRRLSNRNHHSOLTCEIILMESHADAAWPPXEPVNPRLV 1800
Db RESPAAGPRYSEGLSPSKRRRLSNRNHHSOLTCEIILMESHADAAWPPXEPVNPRLV 1800
Qy SGYRRIKNPMDFTMRERLLRGYTSSEEFADALLVFDNCQTFNEDDSEVGKAGHIMR 1860
Db SGYRRIKNPMDFTMRERLLRGYTSSEEFADALLVFDNCQTFNEDDSEVGKAGHIMR 1860
Qy RPFESRWEFFYQKQANL 1878
Db RPFESRWEFFYQKQANL 1878

RESULT 3
ABM85417
ID ABM85417 standard; protein; 1873 AA.

XX AC ABM85417;

XX DT 18-NOV-2004 (first entry)

XX DE Human protein sequence hCP42010.

KW Cytostatic; carcinoma; lymphoma; cancer; human.
XX Homo sapiens.
XX WO2003073826-A2.
XX 12-SEP-2003.
XX 28-FEB-2003; 2003WO-US006235.
XX 01-MAR-2002; 2002US-00087192.
XX (SAGR-) SAGRES DISCOVERY.
XX Morris DW;
XX WPI; 2003-328604/31.
XX Recombinant nucleic acid useful for diagnosis and treatment of carcinoma comprises a nucleotide sequence.
XX Claim 5; SEQ ID NO 666; Opp; English.
XX The present invention relates to novel DNA and protein sequences which are associated with carcinomas. The sequences are useful for: (i) for screening drug candidates; (ii) for screening of bioactive agent capable of binding to Carcinoma Associated Protein (CAP); (iii) for screening of a bioactive agent capable of modulating the activity of CAP; (iv) for evaluating the effect of a candidate carcinoma drug; (v) for diagnosing carcinoma; (vi) for inhibiting the activity of CAP; (ix) as a biochip; (x) for diagnosing carcinoma or a propensity to carcinoma; and (xi) for determining Carcinoma Associated (CA) gene copy number. In addition, the CA genes are useful as DNA vaccines and the CAP are useful as markers of carcinoma including lymphoma. The present sequence is one such CAP. Note: This patent is an equivalent to basic patent US2002182586A1, for which no sequence data was published

XX SQ Sequence 1873 AA;

Query Match 97.6%; Score 9590.5; DB 7; Length 1873;
Best Local Similarity 98.0%; Pred. No. 0;
Matches 1840; Conservative 8; Mismatches 25; Indels 5; Gaps 2;

Qy 1 MEMEANDFHFTGLPPAPAAAGLKPSPSSGGLYTNNGSPMFPPOGKSLNGDVNVNG 60
Db 1 MEM---FANDHFNFTGLPPAPAAAGLKPSPSSGGLYTNNGSPMFPPOGKSLNGDVNVNG 57
Qy 61 LSTVSHTTTSGILNSAPHSSSTSHLHPSVAYDCLMNYSOYPSANPGSNLKDPPLLSQFS 120
Db 58 LSTVSHTTTSGILNSAPHSSSTSHLHPSVAYDCLMNYSOYPSANPGSNLKDPPLLSQFS 117
Qy 121 GGQYPLNGILGGSQSPSPSHNTNLRAGSKQFWANGTHSPMGLNFDQELVDSPPDQNF 180
Db 118 GGQYPLNGILGGSQSPSPSHNTNLRAGSBQFWANGTQSPMGLNFDQELVDSPPDQNF 177
Qy 181 EVCSGIHPDEAAEKEMTVVAENGTGLVCSLLEEXOPELKMCCYNGSVPSVESLHQSVS 240
Db 178 EVGSGIHPDEAAEKEMTVVAENGTGLVLSLELEEEQPELKMCCYNGSVPSVESLHQSVS 237
Qy 241 VLVPDPTVSCLDLDPDHLPDQLEDPTILSEDSLEFPNSLAPPEVSGGLYIGIDDTLMGAED 300
Db 238 VLVPDPTVSCLDLDPDHLPDQLEDPTILSEDSLEFPNSLAPPEVSGGLYIGIDDTLMGAED 297
Qy 301 KLPLXDSFVISALDCPSLNNATAFSLDADDSQTSSTIFASPTSPVILGESVLQNSFDLN 360
Db 298 KLPLEDSFVISALDCPSLNNATAFSLDADDSQTSSTIFASPTSPVILGESVLQNSFDLN 357
Qy 361 NGSDAEQEMETQSSDFPPSLTQAPDQSSNIOQLHATSPAVSPTTSPAVSLVVSPPASP 420
Db 358 NGSDAEQEMETQSSDFPPSLTQAPDQSSNIOQLHATSPAVSPTTSPAVSLVVSPPASP 417
Qy 421 EISPEVCFAASTVVSFAVSVSPASSAVLPAVSLVLEVPFTASVTSKASPTSPASAPPT 480

Db 418 BISPEVCFAASTWSPAVFVSPASSAVLPAVSLVETLTAIVTSKASPTVSPAAREPT 477
Qy 481 ASPANKOVSSFLETTADVEEITGEGLTASGSDVNRRIATPEEVRLLPQHGWRREVRIK 540
Db 478 ASPANKOVSSFLETTADVEEITGEGLTASGSDVNRRIATPEEVRLLPQHGWRREVRIK 537
Qy 541 KGSRWQOETWYGPCGRMKQFPEVIKYLGRNLVHVSRRHFSPSPMPVGDFFPEERDT 600
Db 538 KGSRWQOETWYGPCGRMKQFPEVIKYLGRNVVHVSRRHFSPSPMPVGDFFPEERDT 597
Qy 601 PEGLOWVLSAEETPSRIQATIGKRGRPRNTEKATKEVPKVRGRGPPPKVKITELLNK 660
Db 598 PEGLOWVLSAEETPSRIQATIGKRGRPRNTEKATKEVPKVRGRGPPPKVKITELLNK 657
Qy 661 TDNRPLKLEAQETLNEEDKAKIAKSKKMRQKVORGECLTTIOQARNKKEQETKSLKH 720
Db 658 TDNRPLKLEAQETLNEEDKAKIAKSKKMRQKVORGECLTTIOQARNKKEQETKSLKH 717
Qy 721 KEAKKKSXAEBKGTKEKLEKVKREKKEKVMKEKEEVTAKPACADKTLATQRRLL 780
Db 718 KEAKKKS--KKEKGTKEKLEKVKREKKEKVMKEKEEVTAKPACADKTLATQRRLL 775
Qy 781 EERQKQOILSEMCKPTEDMCLTDHQLPDRSRVPGLTLPSCAFSDCLTIIVEFLHSFGKV 840
Db 776 EERQKQOILSEMCKPTEDMCLTDHQLPDRSRVPGLTLPSCAFSDCLTIIVEFLHSFGKV 835
Qy 841 LGFDPADKDVPSLGVLOEGLLQCGDSLGEVDQLLVRLLKKAALHDGFPSPYCOSLKILGKV 900
Db 836 LGFDPADKDVPSLGVLOEGLLQCGDSLGEVDQLLVRLLKKAALHDGFPSPYCOSLKILGKV 895
Qy 901 SEIPLTRDNVSEIILRCFLMAYGVXPALCDRLRTOPFOAQQPQKAAVLAFVHVLNGSTL 960
Db 896 SEIPLTRDNVSEIILRCFLMAYGVXPALCDRLRTOPFOAQQPQKAAVLAFVHVLNGSTL 955
Qy 961 IINEIDKTLSEMSRVKRNKWIIEGRLRLKTVLAKRTGRSEVEMGRPECLGRRSSRIM 1020
Db 956 IINEIDKTLSEMSRVKRNKWIIEGRLRLKTVLAKRTGRSEVEMGRPECLGRRSSRIM 1015
Qy 1021 EETSMBEEEEESTAAVPGRRGRDGEVDATASSIPELRQIEKLSKROLFFPKKLLHS 1080
Db 1016 EETSMBEEEEESTAAVPGRRGRDGEVDATASSIPELRQIEKLSKROLFFPKKLLHS 1075
Qy 1081 SQMLRAVSLGQDQRRYRRVWLPYLAGIFVEGTEGNLVPEEVIKKEETDSLKVAHAASLNPA 1140
Db 1076 SQMLRAVSLGQDQRRYRRVWLPYLAGIFVEGTEGNLVPEEVIKKEETDSLKVAHAASLNPA 1135
Qy 1141 LFSMKMELAGSNTTASSPARARSRLTKTPGMQPRHPFKSPVRGQDSQPOQAQLQPEAQL 1200
Db 1136 LFSMKMELAGSNTTASSPARARGRPRTKPGSMQPRHLKSPVRGQDSQPOQAQLQPEAQL 1195
Qy 1201 HVPAQOPQLOLQLOSHKGFLEOEGSPLSLQSOHDLQSAPFLSWLSQTSQSHSLSSSV 1260
Db 1196 HAPAQOPQLOLQLOSHKGFLEOEGSPLSLQSOHDLQSAPFLSWLSQTSQSHSLSSSV 1255
Qy 1261 LTPDSSPKGLDPAFSPQPEPEPEDEAESPDLOAFWFNISAQMPCNAAFTPPPAVSEDPQ 1320
Db 1256 LTPDSSPKGLDPAFSPQPEPEPEDEAESPDLOAFWFNISAQMPCNAAFTPPPAVSEDPQ 1315
Qy 1321 TSPSQOLASSKPMNRPSAANPCSPVQFSSSTPLAGLAPKRRAGDPCGEMPOSSTGLGQPKRR 1380
Db 1316 TSPSQOLASSKPMNRPSAANPCSPVQFSSSTPLAGLAPKRRAGDPCGEMPOSSTGLGQPKRR 1375
Qy 1381 GRPSPKPFKQMEQRYLTQLTQAPVPEMCSGWWWIPDPEMLDAMLKALHPRGIREKALHK 1440
Db 1376 GRPSPKPFKQMEQRYLTQLTQAPVPEMCSGWWWIPDPEMLDAMLKALHPRGIREKALHK 1435
Qy 1441 HLNKHRDFLOEVCRLRPSADPIFEPRQLPAFOEGIMSWSPKERTYETDLAVLQWVEELEQR 1500
Db 1436 HLNKHRDFLOEVCRLRPSADPIFEPRQLPAFOEGIMSWSPKERTYETDLAVLQWVEELEQR 1495
Qy 1501 VIMSDLQIRGWTCPSPDSTREDLAYCEHLSQSEBITWRGGRGLAPQRKTTNPDLAV 1560

Db 1496 VIMSDLQIRGWTCPSPDSTREDLAYCEHLSQSEBITWRGGRGLAPQRKTTNPDLAV 1555
Qy 1561 MRLAALQONVRRYLRLPWLPTHEVLEKALLSTPNGAPEGTTTISYEITPRIRWROT 1620
Db 1556 MRLAALQONVRRYLRLPWLPTHEVLEKALLSTPNGAPEGTTTISYEITPRIRWROT 1615
Qy 1621 LQRCRSAHVCLCLGHLERSIAWKSVMKVTCVCRKGDNDDEFLLCQDGRGCHYCHR 1680
Db 1616 LERCRAAQVCLCLGQLERSIAWKSVMKVTCVCRKGDNDDEFLLCQDGRGCHYCHR 1675
Qy 1681 PKMEAVPEGDFCTVCLAQOQVEGSEFTOKPGPKGOKKSGYSLNFSFGDGRRRVLLKG 1740
Db 1676 PKMEAVPEGDFCTVCLAQOQVEGSEFTOKPGPKGOKKSGYSLNFSFGDGRRRVLLKG 1735
Qy 1741 RESPAAGPRYSEERLSPSKRRRLSMRNHSHDLTFCIIILMEMESHDAAWPXPVNPRLV 1800
Db 1736 RESPAAGPRYSEEGSLSPSKRRRLSMRNHSHDLTFCIIILMEMESHDAAWPXPVNPRLV 1795
Qy 1801 SGYRRIIKNPMDFSTMRERLLRGYTSSEEPFAADALLVFNQCOTFNEDDSEVGKAGHMR 1860
Db 1796 SGYRRIIKNPMDFSTMRERLLRGYTSSEEPFAADALLVFNQCOTFNEDDSEVGKAGHMR 1855
Qy 1861 RFFESRWEEFYQGKOANL 1878
Db 1856 RFFESRWEEFYQGKOANL 1873
RESULT 4
AAB95554
ID AAB95554 standard; protein; 1727 AA.
XX AAB95554;
XX AC
XX AC
DT 26-JUN-2001 (first entry)
XX
DE Human protein sequence SEQ ID NO:18183.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
OS Homo sapiens.
XX
FN EPI074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-00116126.
XX
PR 29-JUL-1999; 99JP-00248036.
PR 27-AUG-1999; 99JP-00300253.
PR 11-JAN-2000; 2000JP-00118776.
PR 02-MAY-2000; 2000JP-00183767.
PR 09-JUN-2000; 2000JP-00241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.
PT
PS Claim 8; SEQ ID NO 18183; 2537pp + Sequence Listing; English.
XX
CC The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination

of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention

XX Sequence 1727 AA;

Query Match		88.2%;	Score 8666.5;	DB 4;	Length 1727;
Best Local Similarity		96.7%;	Pred. No. 0;		
Matches 1674;		Conservative 9;	Mismatches 41;	Indels 7;	Gaps 2;
Qy	148	GSQFWANGTHSPMGLNFDSDQELYDSDPDQNFEEYCSGIHPDEAAEKEMTSVAENGTLG	207		
Db	4	GPPSFTSPQSPM-LGSSIQTFAPS-----QEVGSGIHPDEAAEKEMTSVAENGTLG	56		
Qy	208	VCSLEBEXQELKXGCVNGVSPVESLHQEVSVLPDPTVSCLDLDDPSHLDPQLEDTPIL	267		
Db	57	VGSLEBEEQELKXGCVNGVSPVESLHQEVSVLPDPTVSCLDLDDPSHLDPQLEDTPIL	116		
Qy	268	SEDSLEPNLSAPBPVSGGLYGIIDDTLMGAEKDLPLXDSVPI SALDCPSLNNAATAPSL	327		
Db	117	SEDSLEPNLSAPBPVSGGLYGIIDDTLMGAEKDLPLXDSVPI SALDCPSLNNAATAPSL	176		
Qy	328	ADDSQTSTSI PASPTSPVLGESVLQDNSFDLNGSDAEQEBEMETQSSDFPSPSTQAPAD	387		
Db	177	ADDSQTSTSI PASPTSPVLGESVLQDNSFDLNGSDAEQEBEMETQSSDFPSPSTQAPAD	236		
Qy	388	QSSTIQLHPATSPAVPTTSPAVSLVSPAAAPSPISPRVCSPAATVSPVSVSPASS	447		
Db	237	QSSTIQLHPATSPAVPTTSPAVSLVSPAAAPSPISPRVCSPAATVSPVSVSPASS	296		
Qy	448	AVLPAVSLVPLELTASVTSKASPTSPAAAPTASPAKDVSSPLETTADVEEITGEGLT	507		
Db	297	AVLPAVSLVPLELTASVTSKASPTSPAAAPTASPAKDVSSPLETTADVEEITGEGLT	356		
Qy	508	ASGSGDVMMRRRIATPEEVRLPLQHGWRREVRI KKGSHRWQGETWYVYCGKRMKQFPPEVI	567		
Db	357	ASGSGDVMMRCIATPEEVRLPLQHGWRREVRI KKGSHRWQGETWYVYCGKRMKQFPPEVI	416		
Qy	568	KYLSRNVLVSVRRHFSFSPMPVGDFFPEERDTPTEGLQWVLSABEIPSRITQAITGKRGR	627		
Db	417	KYLSRNVLVSVRRHFSFSPMPVGDFFPEERDTPTEGLQWVLSABEIPSRITQAITGKRGR	476		
Qy	628	PRNTEKATKEVPKVGKRGPPKVKITELANKTDNRPLKLEAQETLNEEDKAKIAKSK	687		
Db	477	PRNTEKATKEVPKVGKRGPPKVKITELANKTDNRPLKLEAQETLNEEDKAKIAKSK	536		
Qy	688	KKMRQKVQRGECATTIQOARNKRKQETKSLKHKEAKKKSXAEBKGTQKQELKEKVKR	747		
Db	537	KKMRQKVQRGECATTIQOARNKRKQETKSLKHKEAKKKSXAEBKGTQKQELKEKVKR	596		
Qy	748	EKKKVKMKKEEVTAKPACAKDTLATQRLEERQKQOMILEMMKPTEDMCLTDHQP	807		
Db	597	EKKKVKMKKEEVTAKPACAKDTLATQRLEERQKQOMILEMMKPTEDMCLTDHQP	656		
Qy	808	LPDFSRVPLTLPAGAFSDCLTIVEFLHSFGKVLGDFPAKDVPSLVGLQELLCQGSLSG	867		
Db	657	LPDFSRVPLTLPAGAFSDCLTIVEFLHSFGKVLGDFPAKDVPSLVGLQELLCQGSLSG	716		
Qy	868	EVQDLLVRLKAAALHDPGFPSPYCSQSLKILGKVSSEIPLTRDNVSEILRCFLMAYGVXPAL	927		

Db	717	EVQDLLVRLKAAALHDPGFPSPYCSQSLKILGKVSSEIPLTRDNVSEILRCFLMAYGVXPAL	776		
Qy	928	CDRLRTQFPQAPQOQKAAVLAFVHELNGSTLLINEIDKLTLESMSYRKKNKIIVEGRLR	987		
Db	777	CDRLRTQFPQAPQOQKAAVLAFVHELNGSTLLINEIDKLTLESMSYRKKNKIIVEGRLR	836		
Qy	988	RLKTVLAKRTGRSEVMGRPEECIGRRRRSRIMEETSGMEEEEEEESIAAVPGRGRRDG	1047		
Db	837	RLKTVLAKRTGRSEVMGRPEECIGRRRRSRIMEETSGMEEEEEEESIAAVPGRGRRDG	896		
Qy	1048	EVDATASSIPELERQIEKLSKRQLFFRKLLHSSQMLRAVSLGQDRYRRRYVWLPYLGI	1107		
Db	897	EVDATASSIPELERQIEKLSKRQLFFRKLLHSSQMLRAVSLGQDRYRRRYVWLPYLGI	956		
Qy	1108	FVEGTGELNVPEEVIKKTETSLKVAHAASLNALPSMKMELAGSNNTTASSPARARSPLK	1167		
Db	957	FVEGTGELNVPEEVIKKTETSLKVAHAASLNALPSMKMELAGSNNTTASSPARARSPLK	1016		
Qy	1168	TKPGFMOPRHFKSPVRGQDSQPOAQLOPEAQLHVPAPQPOLOLOLQSHKGFLEQEGSP	1227		
Db	1017	TKPGFMOPRHFKSPVRGQDSQPOAQLOPEAQLHVPAPQPOLOLOLQSHKGFLEQEGSP	1076		
Qy	1228	LSLQSQSHLSQSASFLSWLSQTSQSHSLSSSVLTDPSSPGKLDPPAPSPPEEPDPAE	1287		
Db	1077	LSLQSQSHLSQSASFLSWLSQTSQSHSLSSSVLTDPSSPGKLDPPAPSPPEEPDPAE	1136		
Qy	1288	SSPDLQAFWFI SAQPCNAAPTPLAVSEDOPTPSPOOLASSKPMNRPSAANPCSPVQF	1347		
Db	1137	SSPDLQAFWFI SAQPCNAAPTPLAVSEDOPTPSPOOLASSKPMNRPSAANPCSPVQF	1196		
Qy	1348	SSTPLAGLAPKRRAGDPGEMPGQSTGLQPKRGRPPSKFKQMEQRYLTQLTQAPVPPE	1407		
Db	1197	SSTPLAGLAPKRRAGDPGEMPGQSTGLQPKRGRPPSKFKQMEQRYLTQLTQAPVPPE	1256		
Qy	1408	MCSGMMWIPDPEMLDAMLKALHPRGIREKALHKLHKLHNRDLQEVCLRPSADPIFEPRQL	1467		
Db	1257	MCSGMMWIRDPDPEMLDAMLKALHPRGIREKALHKLHKLHNRDLQEVCLRPSADPIFEPRQL	1316		
Qy	1468	PAFQEGIMSNVSPKKTETDVLAVLQWVEELEFORVIMSDLOIRGWTCPSPDSTREDLAYCE	1527		
Db	1317	PAFQEGIMSNVSPKKTETDVLAVLQWVEELEFORVIMSDLOIRGWTCPSPDSTREDLAYCE	1376		
Qy	1528	HLSDSQSDITWRGPGREGAPQRTTNPLDLAVNRLAALQONKRRYLRPLWPTHEVVL	1587		
Db	1377	HLSDSQSDITWRGPGREGAPQRTTNPLDLAVNRLAALQONKRRYLRPLWPTHEVVL	1436		
Qy	1588	EKALLSTPNGAPEGTTTTEISYEITPRIRIWRQTLQRCRSAHVCLCLGHLERSIAWEKSV	1647		
Db	1437	EKALLSTPNGAPEGTTTTEISYEITPRIRIWRQTLQRCRSAHVCLCLGHLERSIAWEKSV	1496		
Qy	1648	NKVTCLVCRKGDNDDEFILLCDGCDRGCHIYCHRPKMEAVPEGDMFCTVCLAQOVEGETQ	1707		
Db	1497	NKVTCLVCRKGDNDDEFILLCDGCDRGCHIYCHRPKMEAVPEGDMFCTVCLAQOVEGETQ	1556		
Qy	1708	KPGPKGQKRSKYSGLNFSEGDGRRRRVLLKGRSEPAAGPRYSEERLSPKRRRLSRN	1767		
Db	1557	KPGPKGQKRSKYSGLNFSEGDGRRRRVLLKGRSEPAAGPRYSEERLSPKRRRLSRN	1616		
Qy	1768	HHSDLTTCFCEIILMEWESHDAAPFEPVNPRLVSGYRRIIKNPMDFSTMRELLRGGYTS	1827		
Db	1617	HHSDLTTCFCEIILMEWESHDAAPFEPVNPRLVSGYRRIIKNPMDFSTMRELLRGGYTS	1676		
Qy	1828	SEEFADALLVDFNCOTFNEDDSEVGKAGHIMRRFFFSRWEEFYQGKQANL	1878		
Db	1677	SEEFADALLVDFNCOTFNEDDSEVGKAGHIMRRFFFSRWEEFYQGKQANL	1727		

RESULT 5
AAM42025
ID AAM42025 standard; protein; 1589 AA.
XX AC
XX AAM42025;

DT 22-OCT-2001 (first entry)
 XX Human polypeptide SEQ ID NO 6956.
 DE
 XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia.
 XX
 XX Homo sapiens.
 OS
 XX WO200153312-A1.
 PN
 XX 26-JUL-2001.
 PD
 XX
 XX 26-DEC-2000; 2000WO-US034263.
 PF
 XX 23-DEC-1999; 99US-00471275.
 PR 21-JAN-2000; 2000US-00488725.
 PR 25-APR-2000; 2000US-00552317.
 PR 20-JUN-2000; 2000US-00598042.
 PR 19-JUL-2000; 2000US-00620312.
 PR 03-AUG-2000; 2000US-00653450.
 PR 14-SEP-2000; 2000US-00662191.
 PR 19-OCT-2000; 2000US-00693036.
 PR 29-NOV-2000; 2000US-00727344.
 XX
 XX (HYSE-) HYSEQ INC.
 PA
 XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;
 PI Zhou P, Goodrich R, Drmanac RT;
 XX WPI; 2001-442253/47.
 DR N-PSDB; AAI61181.
 XX
 XX Novel nucleic acids and polypeptides, useful for treating disorders such
 PT as central nervous system injuries.
 XX
 XX Example 2; SEQ ID NO 6956; 10078pp; English.
 PS
 XX The invention relates to human nucleic acids (AAI57798-AAI61369) and the
 CC encoded polypeptides (AAM38642-AAM42213) with nootropic,
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: immune system suppression,
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemias and
 CC C.N.S disorders. Note: The sequence data for this patent did not form
 CC part of the printed specification
 XX
 XX Sequence 1589 AA;
 SQ
 Query Match 80.4%; Score 7897; DB 4; Length 1589;
 Best Local Similarity 98.1%; Pred. No. 0;
 Matches 1515; Conservative 7; Mismatches 22; Indels 0; Gaps 0;
 332 QTSTSFASPTSPVVLGSVLQDNSFDLNGSDAEQEEMETQSSDFPPLTQAPADQSSST 391
 Db 1 QTSTSFASPTSPVVLGSVLQDNSFDLNGSDAEQEEMETQSSDFPPLTQAPADQSSST 60
 392 IQLHPATSPVSPPTSPVSLVWSPAAPEISPEVCRAASTVWSPVSVSPASSAVLP 451
 Qy
 Db 61 IQLHPATSPVSPPTSPVSLVWSPAAPEISPEVCRAASTVWSPVSVSPASSAVLP 120

Qy 452 AVSLEVLPTASVTSPKASPVTSAAAAPTASPAKDVSSFLFETTTADVEITGEGLTASGS 511
 Db 121 AVSLEVLPTASVTSPKASPVTSAAAAPTASPAKDVSSFLFETTTADVEITGEGLTASGS 180
 Qy 512 GDMVRRRIATPEEVRLPLQHGWRREVRKKGSHRWQGETWYGPCGKMKKOPPEVVKYLS 571
 Db 181 GDMVRRRIATPEEVRLPLQHGWRREVRKKGSHRWQGETWYGPCGKMKKOPPEVVKYLS 240
 Qy 572 RNLVHSVRRRHFSFSPRMPVGDFFFEERDTPEGLQWVLSAEIIPSRQAITGKGRPRNT 631
 Db 241 RNVVHSVRRRHFSFSPRMPVGDFFFEERDTPEGLQWVLSAEIIPSRQAITGKGRPRNT 300
 Qy 632 EKATKEVPKVRGRGRPPKVIITELLNKTDNRPLKKLEAQETNEEDKAKIAKSKKOR 691
 Db 301 EKATKEVPKVRGRGRPPKVIITELLNKTDNRPLKKLEAQETNEEDKAKIAKSKKOR 360
 Qy 692 QKVORGECLTTIQGAARKRKQETSLKHKEAKKSKAEKEKGTQKSLKEKVKREKKE 751
 Db 361 QKVORGECLTTIQGAARKRKQETSLKHKEAKKSKAEKEKGTQKSLKEKVKREKKE 420
 Qy 752 KVKMKEKEEVTAKAPACKADTKLATQRRLEERQKQMILEEMKKEPTEDMCLTDHQPLPDF 811
 Db 421 KVKMKEKEEVTAKAPACKADTKLATQRRLEERQKQMILEEMKKEPTEDMCLTDHQPLPDF 480
 Qy 812 SRVPGTLPSGAFSDCLTIVEFLHSFGKVLGDFPAKDVPSLGVLOEGLCCGSDSLGEVQD 871
 Db 481 SRVPGTLPSGAFSDCLTIVEFLHSFGKVLGDFPAKDVPSLGVLOEGLCCGSDSLGEVQD 540
 Qy 872 LLVRLAKAALHDPGFPSPYCSQSLKILGKVSIEIPLTRDNVSEIILCFMAYGVXPALCDRL 931
 Db 541 LLVRLAKAALHDPGFPSPYCSQSLKILGKVSIEIPLTRDNVSEIILCFMAYGVXPALCDRL 600
 Qy 932 RTQPFQAPQPOQKAAVLAFPVHELNGSTLIINEIDKLTLESMSYRKNKWIIVEGRLRLKT 991
 Db 601 RTQPFQAPQPOQKAAVLAFVLHELNGSTLIINEIDKLTLESMSYRKNKWIIVEGRLRLKT 660
 Qy 992 VLAKRTGRSEVEMGRPEECLEGRSSRIMEETSGMEBEESIAAIVGRRGRDGEVDA 1051
 Db 661 VLAKRTGRSEVEMGRPEECLEGRSSRIMEETSGMEBEESIAAIVGRRGRDGEVDA 720
 Qy 1052 TASSIPELERQIEKLSKROLFFRKKLHSSOMLRAVSLGODRYRRYRVWVPLAGIFVEG 1111
 Db 721 TASSIPELERQIEKLSKROLFFRKKLHSSOMLRAVSLGODRYRRYRVWVPLAGIFVEG 780
 Qy 1112 TEGNLVPEEVIKKTDSIKVAHAASLNPALFMSMKELAGSNNTASSPARASRPRLKTPG 1171
 Db 781 TEGNLVPEEVIKKTDSIKVAHAASLNPALFMSMKELAGSNNTASSPARASRPRLKTPG 840
 Qy 1172 FMQPRHFKSPVRGQDSQPOQAQLQPEAQLHVPAQPOQLOLQSHKGFLEQSGPLSLG 1231
 Db 841 SMQPHLKSPPVRGQDSQPOQAQLQPEAQLHVPAQPOQLOLQSHKGFLEQSGPLSLG 900
 Qy 1232 QSQHDLSQSAFLSWLSQTSQSHSSLLSSVLTSDSSPKGLDPAQOPPEPEPDEAESPD 1291
 Db 901 QSQHDLSQSAFLSWLSQTSQSHSSLLSSVLTSDSSPKGLDPAQOPPEPEPDEAESPD 960
 Qy 1292 LQAFWFNISAQMPCNAAAPTPELVSEOPTSPQOLASSKPMNRPSSAANPCSPVQFSSTP 1351
 Db 961 PQALWFNISAQMPCNAAAPTPELVSEOPTSPQOLASSKPMNRPSSAANPCSPVQFSSTP 1020
 Qy 1352 LAGLAPKRRAGDPCGEMQSPGTGLQPKRRGRPPSKFFKQMEQRYLTQTAQVPPVEMCSG 1411
 Db 1021 LAGLAPKRRAGDPCGEMQSPGTGLQPKRRGRPPSKFFKQMEQRYLTQTAQVPPVEMCSG 1080
 Qy 1412 WWWTLPDEMLDAMLKALHPGRIREKALKHKLKURDFLOEVLCPSPADPIFEPRQLPAFQ 1471
 Db 1081 WWWTLPDEMLDAMLKALHPGRIREKALKHKLKURDFLOEVLCPSPADPIFEPRQLPAFQ 1140
 Qy 1472 EGINSWSPKEKTYETDLAVLQWVEELRQRYTMSDLQIRGTWCPSPDSTREDLAYCEHLS 1531
 Db 1141 EGINSWSPKEKTYETDLAVLQWVEELRQRYTMSDLQIRGTWCPSPDSTREDLAYCEHLS 1200
 Qy 1532 SQEDITWRGPGREGLAFQRTTNPLDLAVMRLAALAEQNVKRYLREPLWPTHEVVLEKAL 1591

[illegible]

RESULT 6

REC'D: 9
ABM85416

ID ABM85416

24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857

AC ABM85416;

XXXXXX

DT 18-NOV-2004 (first entry)

XX XX

DE Mouse protein se

XX

KW Cytostatic; carcinoma; lympho

XX

OS Mus musculus.

XX

PN WO2003073826-A2.

[illegible]PD
yy
12-SEP-2003.

XX
3000-6848-8C

PF
XX
28-FEB-2003; 2003WO-US0006235.

XX
PR 01-MAR-2003: 2002US-00087192

FR 01-MAR-2002; 200205-0008/192.
XX

PA (SAGR-) SAGRES DISCOVERY.

FA (SAGR-) SAGRES DISCOVERY.
XX
XX

PI Morris DW:

1
3
5
2
4
4
4
0
1

DR WPI; 2003-328604/31.

XX
XX

PT Recombinant nucleic acid useful for diagnosis and treatment of carcinoma

PT comprises a nucleotide sequence.

XX
XX

PS Claim 5; SEQ ID NO 663; Opp; English.

XX

CC The present inven

Query Match 69.0%; Score 6777.5; DB 7; Length 1586;
Best Local Similarity 69.3%; Pred. No. 0;
Matches 1323; Conservative 81; Mismatches 128; Indels 377; Gaps 9;

Qy	1	MEWEANEANDHFNFTGLPPAPAAAGLKPSPSSGEGLYTNGSPMNFPOOGKSLNGDNNVNG	60
Db	24	MEM- --EANDHFNFTGLPPAPAAAGLKPSPSSGEGLYTNGSPMNFPOOGKSLNGDNNVNG	80
Qy	61	LSTVSHTTTSGILNSAPHSSSTSHLHHPSVAYDCLWNYSQYPSANPGSNLKDPLLQSFS	120
Db	81	LSTVSHTTTSGILNSAPHSSSTSHLHHPNVAYDCLWNYSQYPSANPGNNLKDPLLQSPF	140
Qy	121	GGQYPLNGILGGQRQSPSPSHNTNLRAGSKQFWANGTHSPMGLNFDQSQELYDFFPDQNF	180
Db	141	GGQYPLNGILGGNRQSPSPSHNTNLRAGSQEFWANGTQSPMGLNFDQSQELYDFFPDQNF	200
Qy	181	-----EVCSGHPDEAAEKEMTSVVAENGTLGVCS	210
Db	201	VMPNGPPSPFTSQTPSMLGSSIQTFAPSDQVSDIHPDEAAEKELTSVVAENGTLGVCS	260
Qy	211	LELEEXPELKMCGYNGSPSVESLHQEVSVLPDPTVSCLDDBPSHLDOLEDTPILS	270
Db	261	LELEEQPELKMCGYNGSPSVESLHQEVSVLPDPTVSCLDDBPSHLDOLEDTPILS	320
Qy	271	SLPEPFNSL- APEPVSGLYGIDDTLWGABDKLPLXDSPVISALDCPSLNNATAFSLAD	329
Db	321	SLPEPDSLAAAPEPVSGLYGIDDAELWGABDKLPLEGNPVISALDCPALSNANAFSLAD	380
Qy	330	DSQTSSTIFASPTSPPVLGESVLQDNSFDLNNGSDAEQEEMETQSSDFFPSLTQPPADQS	389
Db	381	DSQTSASIFVSPSTSPPVLGESVLQ-----	404
Qy	390	STIQLHPATSPAVSPSTTSPAIVLWVSPAASPEISPEVCPAAATVWSPAVFVWSPASSAV	449
Db	405	-----	404
Qy	450	LPAVSLVPLUTASVTSKASPVTSPPAAFTSPANKDVSSFLIETIADVEIITGEGLTAS	509
Db	405	-----GSPSPSPAAAFQTVSPARKNVSSAPKARADRETTGGAVAS	446
Qy	510	GSQDVMBRRLATPEEVRLPLQHGWRREVRILKKGSHRWQGETWYVGPCGKMKQFPEVIKY	569
Db	447	GSQDVLMKRRLATPEEVRLPLQHGWRREVRILKKGSHRWQGETWYVGPCGKMKQFPEVIKY	506
Qy	570	LSRNLVHSVRREHFSRPMVPVGDFFPEERDTPGLQWVLSABEISPSIOAITGKRGRPR	629
Db	507	LSRNVVHSVRREHFSRPMVPVGDFFPEERDTPGLQWVLSABEISPSIOAITGKRGRPR	566
Qy	630	NTEKATKEVPVKRGRRPPVKIITELLNKTNRPLKLEAQETLNEEDKAKIAKSKKK	689
Db	567	NNEKAKNKEVPVKRGRRPPKIMPELLNKTNRPLKLETOIILEDKAKMTKNKKK	626
Qy	690	MROKVORGECLTTITQOARNKRQETKSLKHKAKKKXAEKEKGTKQBLKEKVKREK	749
Db	627	MROKVORGESQTPVQOARNKRQDKTSLKQKOTKKK-----	663
Qy	750	KEVKMKKEBEVTKAPACKADKTLATORLREBRQQMILEEMKKPTEDMCLTDHQPLP	809
Db	664	-----LKRRLEEQRQQAILEEMKKPTEDMCLSDHQPLP	697
Qy	810	DFSRVPGLTILPSCAFSDCLTIVEFLHSFGKVLGFDPAKDVPSLGVLOEGLLCOGDSLGEV	869
Db	698	DFTRIPEGLTSSRAFSCLTIVEFLHSFGKVLGFDLTQDVPSLGVLOEGLLCOGDSLDKV	757
Qy	870	QDLLVRLKLAALHDPGPSYCSQLKILGEKVSIPILTRDNVSBILRCFLMAYGVXPALCD	929
Db	758	QDLLVRLKLAALHDPGLPPYCSQLKILGEKWSIPILTRDNVSBILRCFLMAYRVSPFCD	817
Qy	930	RLRTQPFQAPPOOKAAVATAPVHELNGSTLIINEIDKTLBSMSSYKRNKWIIVEGRRLRL	989
Db	818	SLRTQPFQAPPOOKAAIATLAFVHELNSPTIINEIDKTLBSYSSCRKNKWIIVEGRRLRL	877

QY	217	QPELKMCGYNGSVPSVESLHQEVSVLVPDPTVSCLDDPSHLPDQLEDTPILSEDSLEPFN	276
Db	242	QPELKMCGYNGSVPSVESLHQEVSVLVPDPTVSCLDDPSHLPDQLEDTPILSEDSLEPFN	301
QY	277	SLAPEPVGGLYGIDDTLMAEDKLPXDSPVISALDCPSLNATATFSLADDSQTSTS	336
Db	302	SLAPEPVGGLYGIDDTLMAEDKLPLEDSPVISALDCPSLNATATFSLADDSQTSTS	361
QY	337	IFASPTSPVVLGESVLQDNSFDLNGSDAEQEBEETQSSDDPPSLTQAPDQSSSTIQLHP	396
Db	362	IFASPTSPVVLGESVLQDNSFDLNGSDAEQEBEETQSSDDPPSLTQAPDQSSSTIQLHP	421
QY	397	ATSPAVSPTTSPAVSLVSPASPEISPEVCPAASTVVSFAVSVSPASSAVLPVAVSLE	456
Db	422	ATSPAVSPTTSPAVSLVSPASPEISPEVCPAASTVVSFAVSVSPASSAVLPVAVSLE	481
QY	457	VPLTASVTSKSPVTSAAAAPPASPANKOVSSFFLETTADVBEITGEGLTASGSDVMR	516
Db	482	VPLTASVTSKSPVTSAAAAPPASPANKOVSSFFLETTADVBEITGEGLTASGSDVMR	541
QY	517	RRIATPEEVRLPLQHGWRREVRIKKGSHRWQGETWYVYPCGRMKQFPFVVKYLSRNLVH	576
Db	542	RRIATPEEVRLPLQHGWRREVRIKKGSHRWQGETWYVYPCGRMKQFPFVVKYLSRNLVH	601
QY	577	SVRREHFSFPRMPVGDFFEBRDTPEGLQWVLSAEEIPSRITQAITGKRGPRNTEKAKT	636
Db	602	SVRREHFSFPRMPVGDFFEBRDTPEGLQWVLSAEEIPSRITQAITGKRGPRNTEKAKT	661
QY	637	KEVPKVRGGRGPPKVKITELLNKTNDNRLPKLEAQETLNEEDKAKIAKSKKQVRQVQR	696
Db	662	KEVPKVRGGRGPPKVKITELLNKTNDNRLPKLEAQETLNEEDKAKIAKSKKQVRQVQR	721
QY	697	GECLTTIQOARNRKEQETKSLKHKEAKKSKYAEKKGKTKOEKLKEKVYKEKKEKVK	754
Db	722	GEQTTIQOARNRKEQETKSLKHKEAKKSKYAEKKGKTKOEKLKEKVYKEKKEKVK	779
RESULT 8			
ADRO9529			
ID	ADRO9529	standard; protein; 524 AA.	
XX	AC	ADRO9529;	
XX	DT	04-NOV-2004 (first entry)	
XX	DE	Human protein useful for treating neurological disease Seq 3035.	
XX	KW	human; oligo-capping method; diagnostic marker; gene therapy;	
XX	KW	osteoporosis; neurological disease; Alzheimer's disease;	
XX	KW	Parkinson's disease; dementia; short memory; cancer;	
XX	KW	sense or motor function; emotional reaction; fear response; panic;	
XX	KW	osteopathic; neuroprotective; nootropic; antiparkinsonian; cytostatic;	
XX	KW	tranquilliser.	
OS		Homo sapiens.	
XX	XX		
PN	EP1447413-A2.		
XX	PD	18-AUG-2004.	
XX	PF	12-FEB-2004; 2004EP-00003145.	
XX	PR	14-FEB-2003; 2003JP-00102207.	
PR	09-MAY-2003; 2003JP-00131452.		
XX	XX	(REAS-) RES ASSOC BIOTECHNOLOGY.	
XX	XX	Isogai T, Yamamoto J, Nishikawa T, Isono Y, Sugiyama T, Otsuki T;	
PI	Wakamatsu A, Ishii S, Nagai K, Irie R,		
XX	DR	WPI; 2004-583265/57.	
DR	N-PSDB; ADR07573.		
XX	XX		

PT	New 1995 cDNA, useful for treating osteoporosis, neurological diseases, Alzheimer's diseases, Parkinson's diseases, dementia and various cancers.	
XX	Claim 1; SEQ ID NO 3035; 2686pp; English.	
XX	This invention relates to novel, isolated full length human cDNA molecules and the encoded proteins thereof. Specifically, it refers to cDNA clones obtained by an oligo-capping method, where none of these clones are identical to any known human mRNAs. The present invention describes an immunoassay to identify agonists and antagonists, as well as antibodies, antisense molecules and siRNAs that can all be used to bind to and modulate expression of the cDNA molecules. As such, these molecules are useful for diagnostic markers or therapeutic targets for the various diseases or morbid states. In particular, they are useful in gene therapy for treating osteoporosis, neurological disease, Alzheimer's disease, Parkinson's disease, dementia, short memory and various cancers, as well as for maintaining equilibrium of sense or motor function, and for treating emotional reaction, fear response and panic. Accordingly, they exhibit osteopathic, neuroprotective, nootropic, antiparkinsonian, cytostatic and tranquilliser activities. This polypeptide is a protein encoded by a full length human cDNA sequence of the invention. NOTE: This sequence is not given in the sequence listing of the specification but can be obtained on CD-ROM from the European Patent Office, Vienna Sub-office.	
XX	Sequence 524 AA;	
SQ		
Query Match	28.1%; Score 2766; DB 8; Length 524;	
Best Local Similarity	94.0%; Pred. No. 4.1e-177;	
Matches	513; Conservative 5; Mismatches 6; Indels 22; Gaps 1;	
QY	1333 MNRESAANPCSPVQFSSTPLAGLAPKRRAGDPGEMPOSPTGLGQPKRRGPPSPKFFQME	1392
Db	1 MNRESAANPCSPVQFSSTPLAGLAPKRRAGDPGEMPOSPTGLGQPKRRGPPSPKFFQME	50
QY	1393 QRYLTQLTAQFVPPPEMCGSWMWIPDPEMLDAMLKALHPRGIREKALHKLHNRDFLQEV	1452
Db	51 -----PPKMCSGMWIIIRDPEMLDAMLKALHPRGIREKALHKLHNRDFLQEV	98
QY	1453 CLRPSADPIPEPRLPAPFQEGIMSWSPKETYETDLAVLQWVEELEQVIMSDLIQIRGWT	1512
Db	99 CLRPSADPIPEPRLPAPFQEGIMSWSPKETYETDLAVLQWVEELEQVIMSDLIQIRGWT	158
QY	1513 CPSPDSTREDLAYCEHLSDSQEDITWRGPGREGLAPQRTKNPLDLAVMLAALAEQNVKR	1572
Db	159 CPSPDSTREDLAYCEHLSDSQEDITWRGPGREGLAPQRTKNPLDLAVMLAALAEQNVKR	218
QY	1573 RYLREPLWPTHEVVLKALLSTPNGAPEGTTTETISYETTPRIRWQTLQCRSAAHVCL	1632
Db	219 RYLREPLWPTHEVVLKALLSTPNGAPEGTTTETISYETTPRIRWQTLQCRSAAHVCL	278
QY	1633 CLGHLERSIAWEKSVNKTCLVCRKGNDEFFLLLCDGDCRGCHYICHRPKMEAVPEGDFW	1692
Db	279 CLGHLERSIAWEKSVNKTCLVCRKGNDEFFLLLCDGDCRGCHYICHRPKMEAVPEGDFW	338
QY	1693 CTVCIAQVGEFTQKPGPKRGOKRKGSLNFSQDGRRRRLVLLKGRSPAAGPRYSE	1752
Db	339 CTVCIAQVGEFTQKPGPKRGOKRKGSLNFSQDGRRRRLVLLKGRSPAAGPRYSE	398
QY	1753 ERLSPSRRLRLSMNRHSHDLTFCEIILMEMESHDAAMPFEPVNPRLVSGYRRIRIKNPM	1812
Db	399 ERLSPSRRLRLSMNRHSHDLTFCEIILMEMESHDAAMPFEPVNPRLVSGYRRIRIKNPM	458
QY	1813 FSTWRERLLRGYTSSEFEAADALLVFDNCOTFNEDDSEVKGAKHIMRRFPESRWEFPYQ	1872
Db	459 FSTWRERLLRGYTSSEFEAADALLVFDNCOTFNEDDSEVKGAKHIMRRFPESRWEFPYQ	518
QY	1873 GKQANL 1878	
Db	519 GKQANL 524	
RESULT 9		

Transcriptional regulator gene family containing bromodomain - may be expressed in testis tissue and is useful for treatment of cancer and other proliferative disorders.

xx PS Claim 1; Page 100-116; 187pp; Japanese.

xx CC This sequence represents the human BAZ2-beta protein, a member of a

xx CC family of transcriptional regulator genes containing a bromodomain (BAZ,

xx CC Bromodomain with Atypical Zinc finger) which are expressed specifically

xx CC in testis tissue and also in certain tumour lines. Transgenic cells may

xx CC be used for the preparation of the BAZ1-alpha, BAZ2-alpha and

xx CC BAZ2-beta proteins. These proteins can be used in the treatment of cancer

xx CC and other proliferative disorders, and in screening of compounds for

xx CC their binding ability to the expression products (e.g. for use as drugs

xx CC by modulation of transcriptional regulation)

xx SQ Sequence 1972 AA;

Query Match 25.3%; Score 2483.5; DB 2; Length 1972;
 Best Local Similarity 32.2%; Pred. No. 3.2e-157;
 Matches 680; Conservative 329; Mismatches 678; Indels 426; Gaps 68;

QY 1 MEMEANEANDHFNFTGLPPAASAGLKPSPS-----SAGELYTNGSPMFPQQ 48

DB 43 MESSNSDSD-----SGTSDTSSEGISSDSDDDLEDEEEDDQSEEDSDSESAQ 98

QY 49 GKSLGDNVNVNGLSTVSHSTTSGILNSAPHSSSTSHLHP--SVAYDCLWNYSQVPSANPG 107

DB 99 HKS-NNQVLLHG---ISDPKADG--QKATEKAQEKRIHQPLPLAFE-----SQTHSFQ-- 145

QY 108 SNLKDPPLLSQ-----PSGGQVPLNGI-----LGSROPSS----- 138

DB 146 SQQKQPVLSQOLPPIFOSSQAKESVNNKHTSVIQSTGLSVNVKPLSLVNOAKKETVMKL 205

QY 139 --PSHNTNLRAGSQKFWANGTHSPMGLNFDQELYDSPDO-----NFEVCSGI-- 186

DB 206 IVPSPDV--LKAGNKMTSESSLTSELKSKREQYKQAPPSQKQKQESSKSLKVKVLAALSN 264

QY 187 -----HPDEAAEKEMTSVAENGTLVCSLEEXQPELKMCGYNGSVPSVLSHQ 237

DB 265 PKATSSSPAHPKQTLNHNHPFLTNAL-----LGNHQP-----NG---VIQSVIQ 307

QY 238 EVSVLVPPDPTVSCLDPSHLDPQLEDTPILSEDSLEPPNSLAPVPVSGGLYIGDIDTEL-- 296

DB 308 EAPL-----ALTTKTKMQSKINEN--IAAASSTPFSFVNLSTSGRRTPFGNQTPVMP 357

QY 297 -----GAEDKLPLXDSPVISALDCPSLNNATAFSLDADDSTSTSIASPTSPVPL 347

DB 358 SASPLHSQGEKAVSNVNPVKI-----QHSHPAKSLVGEFRGTDSDIFSSKDSN 412

QY 348 GESVLQDNSFDLNGSDAEQEMETQSSDFPPLSTQPA-----PDQSSTIQLHPAT 398

DB 413 EDEEEDDEEEDDEDDSDSQSESDSNSESOTEGSEEDDDDDKQDESDSDTEGEKT 472

QY 399 SPAPVPTTSPAVSLVSPAPSPETSPEVCPAASTVSPAVTSVSPASSAVLPAVSLVP 458

DB 473 SMKLNKTT-----SVKSPSMS----- 489

QY 459 LTASVTSPKASPVTSAPAAFTASPAKNDVSSFLETTADVETTCGELTAS--GSGDVMR 517

DB 490 LTGHSTPNLNLHAKAPGA--PAALCSQSQAFLGTSSST-----LTSSPHSGTSKRR 541

QY 518 RIATPEEVRLPLQHGWRREVRKKGSHRWQGETWYIGPCGRMKQFPVPIKYLGRNLVHS 577

DB 542 RVTDERELRIPLYGQWQRETRIRNPGGLQGEVAYAPCGKKLRQYPEVIKYLGRNGIMD 601

QY 578 VRRHFSFSPMPVGDPEERDTPBGLQWVLSAEIISRIOATGKGRGRPNTEKATK 637

DB 602 ISRONFSPSARIVGDFEARDGPOEMQWCLLKEEDVIPRIAMEGRGRPNPDQRAR 661

QY 638 EVPKVRGRGPPKVKITELNKTNDNRPLKLEAQE-----TLNEEDKAKTAK 685

DB 662 EESMRRKGRPPNVNDAEFLDNADAKLRLKIQAEIARQAAQIKLRLKQKQSOARVAK 721

QY 686 SKKQMR-----QKVQRGECLTITQGAQNRKRQETKSLKH-----KEAKKKSXAEEK 733

DB 722 EAKQQAAMAAEKKRQKEQIKIMKQOEKIKRIQQIRMEKELRAQQIILEAKKKKKEAAN 781

QY 734 GK--TKQEKLEK-----VKREKKEVKM-----KEKEVTKAKPACKAD 771

DB 782 AKLAEAKRIIEKEMRRQAVLLKHQERRRRQHMMLMKAWEARKKAEKBERLKQEKDE 841

QY 772 KTLATQRRLEERQKQMILEMKKPTEDMCLTQHPIPLPDFSRVPGLTLPSCAFSDCLTIV 831

DB 842 KRLNKERKLEORRLELEMAKELKAPNEDMCLADQKPIPELPRIPLGLVLSGTSFSDCLMVV 901

QY 832 EFLHSFGVLGDFPAKDVPSLGLVLEQGLCOGDSLGEVQDILLVRLKLAALHDPGPSPCYQ 891

DB 902 QFLNFGKVLGFDVNIIDVNLVLEGLNLGDSMGEVQDILLVRLLSAAVCDPGLITGYK 961

QY 892 SLKLILGEKVSIEPIPLTRDNVSEILRCFLMAYGVXALCDRLRTPQFAQPPQKAAVLAFP 951

DB 962 AKTALGEHLNVGNVNRDNVSEILQIFMEAHGCGQTELTESLTKAFQAHTPAKASVLAFL 1021

QY 952 VHELNGSTLIINEIDKTLSESMYSYKKNKIIVEGRRLRLKTVLAKTRGRSE----VEMGRP 1007

DB 1022 INELACSKSVSEIDKIDYMSNLRDKVWVEGKRLKRIIHAKTGKRTSGGIDIGEE 1081

QY 1008 EECILG-----RRRSRIME-----ETSGMEEEEEEESIAAEVGRRR--RDGEV 1049

DB 1082 QHPLGTPTGRRKRRKGGSDYDDDDDDSDQDDEDEDEDEKDGKKTDCEDDEDE 1141

QY 1050 DATASSIPELEROIEKLSKROLFRKLLHSSQMLRAVSLGQDRYRRYRWLPYLAGIFV 1109

DB 1142 GDQAASVEELEKQIEKLSKQSQYRRKLFDAHSLRSVMFGPDRYRRYRWLPYRCGGIFV 1201

QY 1110 EGTENLVPEVVIKKTDSLKVAHAASLNPLAFSMKMKELAGSNNTASSPARARS--PLKT 1168

DB 1202 EGMESGEGLEB--IAKERELKKAESVQKEMP-----ETSGDSLNCSTNDCEQKEDLKE 1256

QY 1169 KPG---FMQ--PRHFKSPVRGQDSQEQQAQLOPEAQLHVPAQPPQLQLOLQSHKGFLEQ 1224

DB 1257 KDNTNLFQKPGSPSKLSKLE---VAKMPPESEVMT-----KPNAGAN 1298

QY 1225 GSPSLSQS--QHDLS--QSAPLSWLSQTSQSHSSLLSSVL--TPDSSPGKL--DPAPSPP 1278

DB 1299 GCTLISYQNSGKHSLSGVSQ-----TATQSVNEKADSNLNTGSSGKGYSLPND-- 1350

QY 1279 EEPEDBAESPDLOAFWFNISAQPCN-----AAPTPLAVSDEQPTTSPQQ 1326

DB 1351 -----QLKLTLTEKRNQWFSLLPTPCDDTSLTHADMSTASLVTPOSPPSKSPPTAP 1405

QY 1327 LASSKPMNRPSAAMP--CSPVQF--SSTPLAGLAKRRRAGDPG-----EMPOSPT 1372

DB 1406 LGSS--AQNPVGLNPFALSPLOVKGVGVMGL---QFCGWPTGVVTSNIPFTLSVPSLGS 1460

QY 1373 GLGQPKRGR-----PPSKF---FKQMEQRYLTOLTA-----QPVPPEMCS 1410

DB 1461 GLGISEGNNGSFLTSNVASSKSPVPQNEKATSAQPAAVEVAKVPDPFPPSPKPIPEEQF 1520

QY 1411 GWWIIPPEMLDAMLKALHPRGIREKALHKLHNRDFLOVCLRPSADPIFEPRQLPAF 1470

DB 1521 GWWRIIDPEDLKALLKVLHLRGIREKALQKIQKHLDYITQACLKNKQDAIIELENEEN 1580

QY 1471 Q---EGIMSWSPKBTYETDLAVLQWVEELEQRVIMSDLQIRGWTCTSPDSTREDLAYCE 1527

DB 1581 QVTRDIVENWSVEQAMMDLSVLQVQVEDLERRVASASLQVKGWMCPEPASEREDLVYFE 1640

QY 1528 HLSDS-----QEDITWRGPGREGL--APOKITNPLDLAMRLAALAEQNVKRYLREPLWPT 1582

DB 1641 HKSFYKLCHEHGFTEDESSAHLERKSNPNLDIAVTRLADLERNIERR----- 1691

QY 1583 HEVLEKALLSTPNGAPEGTTTETSYETTPRIRWROTQRCSAAHVCCLCGLHLSRIA 1642

DB 1692 -----TEEDIAFGLRVRRALSEARSAQAQVALLCIQLOKQIA 1728

QY 1643 WEKSVNKKVTCVLCRKGNDPEFLLCDCGDCRCHTYCHRPKNMEAVPEGDWFTVCVLAQ--- 1699

DB 1729 WEKSIKMYVQICRKGNDNEELLLCDGCDKCHTYCHRPKITTTIPDGDWFCPACIAKASG 1788

Qy	1700	-----QVEGFTQTQKPGPKQGQK-----RKSGVSLNFSFEGDGRRRRVLLK	1739
Dd	1789	QTLIKLKLHVKGKKTNF---SKGKKVTLTGTDDEDSASTSSSLKFGNKDLQKRKM---	1842
Qy	1740	GRESPAAGPRYSERLSPSRRLRLSMENHSDLTFCETILMESHDAAMPXEPVNPRL	1799
Dd	1843	-ENTSINLSKQSGFTSVKPKP----RDDSKDLALCSMLITEMETHEDAWPFLLPVNLKL	1897
Qy	1800	VSGYRIIRIKNPMDFSTWRERLLRGGYTSSBEFAADALLVFDNCOTFNEDDSEYVKAGHIM	1859
Dd	1898	VPGVKKVIKPMDFSTIREKLSQQYPNLETFFALDVRLVFDNCETFNEDDDSIGRACHNM	1957
Qy	1860	RPFESRWEEFYQ	1872
Dd	1958	RKYFEKKWTDTFK	1970
 RESULT 11 ABR64241 ID ABR64241 standard; protein; 1972 AA.			
XX	AC	ABR64241;	
XX	DT	15-OCT-2003 (first entry)	
XX	XX	Angiogenesis protein BNO150.	
DE	XX	Cytostatic; antirheumatic; antiarthritic; antidiabetic; ophthalmological;	
KW	KW	antipsoriatic; antiarteriosclerotic; cardiant; vasotropic; angiogenesis;	
KW	KW	gene therapy; vasculature; cancer; rheumatoid arthritis; psoriasis;	
KW	KW	diabetic retinopathy; cardiovascular disease; atherosclerosis;	
KW	KW	ischemic limb disease; coronary artery disease.	
XX	OS	Homo sapiens.	
PN	XX	WO2003027285-A1.	
XX	XX	03-APR-2003.	
PD	XX		
XX	XX	19-SEP-2002; 2002WO-AU001282.	
PF	XX		
XX	XX	27-SEP-2001; 2001AU-00007973.	
PR	PR	27-SEP-2001; 2001AU-00007974.	
PR	PR	11-OCT-2001; 2001AU-00008210.	
PR	PR	23-OCT-2001; 2001AU-00008532.	
PR	PR	13-NOV-2001; 2001AU-00008838.	
PR	PR	28-AUG-2002; 2002AU-00951032.	
XX	XX	(BION-) BIONOMICS LTD.	
PA	XX		
PI	XX	Gamble JR, Hahn CN, Vadas MA;	
PT	XX	WPI; 2003-354655/33.	
DR	DR	N-PSDB; ACF34516.	
XX	XX		
PT	PT	New angiogenic genes and polypeptides, useful for diagnosing,	
PT	PT	prognosticating or treating an angiogenesis-related disorder, e.g.	
PT	PT	cancer, rheumatoid arthritis, diabetic retinopathy, psoriasis or	
PT	PT	cardiovascular diseases.	
XX	XX		
PS	PS	Claim 15; SEQ ID NO 176; 90pp; English.	
XX	XX		
CC	CC	The invention relates to the isolation of novel genes (ACF34446-ACF34559) encoding proteins (ABR64180-ABR64281) involved in the processes of angiogenesis. The nucleic acid molecules are useful in identifying and/or obtaining full-length human genes involved in an angiogenic process. The nucleic acid molecule, polypeptides or complexes encoded, cells or genetically modified non-human animals derived from these are useful for the screening of candidate pharmaceutical compounds used in treating angiogenesis-related disorders. They are also useful for diagnosing, prognosticating or treating an angiogenesis-related disorder, which involves uncontrolled or enhanced angiogenesis or is a disorder in which	

Db	782	AKLEAEKRIKEMRRQAVLLKHQBERRRRQHMLMKAMEARKKAEERKLQKQRDE	841
Qy	772	KTLATORLERQKQOMILEMKKPTEDMCLTDHOPLPDFSRVPGLTLPSGAFSDCLTIV	831
Db	842	KRLNKEKTLQRRLEMAKELKPNEDMCLAQPLPELPRIPGLVLSGSTFSDCLMVV	901
Qy	832	EFLHSFKVLGFDPKADVPISLVGLQEGLLCQGSGLGEVDLLVRLLLKAAHLDPGFPSYCQ	891
Db	902	QFLRNFQKVLGFVDNIDVPNLSVLQEGLLNIGDSMGEVDLLVRLLSAAVCDPGLITGYK	961
Qy	892	SLKILGKVBSEIPLTRDNVSEIILRCFLMAYGVVPALCDRLRTOFPQAPPOQKAAVLAFP	951
Db	962	AKTALGELLNMGVNRDNVSEIILQIFMEAHCGGOTELTSIKTFAOHTPAQKASVLAFL	1021
Qy	952	VHELNGSTLIINEIDKTLSEMSVYRKNKWTVEGRLKELTKVLAKRTGRSE----	1007
Db	1022	INELAKSVSEIDKNIDYMSNURDKWVEGRLKRLRIHAKTKGRITSGDIDGEE	1081
Qy	1008	EECLG-----RRSSRIME-----ETSGMEEEEEESIAAVPGRGR--RDGEV	1049
Db	1082	QHPLTPTPKRRRKGGSDYDDDDDDDDQDDEDEEDKEDKGKKTIDICEDEDE	1141
Qy	1050	DATASSIPELERQIEKLSKQOLFPRFKLLHSQMLRAVSGQDRYRRRYWVLYLAGIFV	1109
Db	1142	GDQAAASVEELEKQIEKLSKQOSQYRKLFPASHLSRSMFQDPYRRRYWVILPRCGGIFV	1201
Qy	1110	EGTEGNLVPBEVIKKTDSLKVAAHASLNALPFSMKMELAGSNNTASSPARASR-PLKT	1168
Db	1202	EGMESGGLSE-TAKEREKUKKXSAESVOIKBEFM-----ETSGDSLNCNTDHCQKEDLKE	1256
Qy	1169	KPG---FMQ--PRHFKSVPVRGDSBPQAOQLQBPQAQLHVPAPQPOLQLQSHKGFLQE	1224
Db	1257	KDNTNLFQKPGSFKLSKLL-----VAKMPPSEVMTP-----KPNAGAN	1298
Qy	1225	GSPLSLGQS-OHDL--QSAFLSWLSQTSKSHLSLLSSVL--TPDSSPGKL-DPAPSQPP	1278
Db	1299	GCTLSYQNSGKSHLSGVQS-----TATQSNVEKADSNLNTGSSGPGKFFYSPLPND--	1350
Qy	1279	EEPEPDABEASPDLOAFWFNISAMQPCN-----AATPPPLAVSEDOPTSPQ	1326
Db	1351	-----QULLKLTQRNQWFSLLPRTPCDDTSLTHADMTASLVTPOSPFKSPSPFAP	1405
Qy	1327	LASKPMNRPSAAMP--CSPVQF--SSTPLAGLAPKRRAGDPG-----EMPOSPT	1372
Db	1406	LGSS--AQNVPGLNPFALSLOYKGVSMGL---QFCGFTGVVTSNIPTLSVPSLGS	1460
Qy	1373	GLGPKRRGR-----PPSKF---FKOMBQRYLTQTA-----QVPVPMCS	1410
Db	1461	GLGLESGNGSFLTNSVASSKSESPVQNEKATSAQPAAVEKVPDFPSPKPIPEMQF	1520
Qy	1411	GWMIIPPEMLDKALHPRGIREKALHKLHNKHRDFOEVCCLRPSADIFEPRLQPAF	1470
Db	1521	GWRIIDPEDKALLKVLHLURIREKALQKQIQKHLDYITQACLKNKDVAIIELNEEN	1580
Qy	1471	Q-----EGTMSWSPKEKTYETDLAVLQWVEELEORVIMSDQIRGWTCPSPDSTREDLAYCE	1527
Db	1581	QVTRDIIVENWSVEQAMENDLSVLQOVEDLERRVASASLQVGMWCEPEASEREDLYFE	1640
Qy	1528	HLSDS-----QEDITWRGFEGL-APQRKTNPLDLAVMRLAALQONVKRYLREPLUMPT	1582
Db	1641	HKSPFKCKEHGDGEFTGEDESSAHLERKSDNPLDIAVTRLADLERNIERR-----	1691
Qy	1583	HEVVLEKALLSTPNGAPEGITTELSYBITPRIRWROTQORCSAAHVCLGLHLEBSIA	1642
Db	1692	-----LEEDIAPLRVWRRALSEARSAAQVLCIQOQKXSIA	1728
Qy	1643	WEKSNVKTCLVCRKGNDFELLCDGCDRGCHYIHRPKMAEVPEDWFTCVCLAQ---	1699
Db	1729	WEKSIMKYCOICRKGNEELLILLCDGCKGCHYIHRPKITTTIPDGWFCPACIAKASG	1788
Qy	1700	-----QVEGFTQKPGFPKRGOK-----RKSGVSLNFSBGDRRRRVLLK	1739

Db	1789	QTLKIKLHVKKKWTNE---SKGKKVTLTGDTEDDSASTSSSLKRGKNDLQKRM---	1844
Qy	1740	GRSPAAQPRYSBERLSPGKRRRLSRNRNHSDLTFCFIIMEMESHDAWPFXPENPRL	1799
Db	1843	-EENTSINLSKQBSFTSVKKPK---RDDSKDLALCSMLITETWETHEDAWPFLPVNLK	1897
Qy	1800	VSGYRIIKNPMDFSTWRERLLRGVTSSEEFADALLVPDNCQTNEDDSEVKGACHIM	1859
Db	1898	VPGYKKVKKPMDFSTIREKLSGQYENLETFALDVLRLVPDNCETFNEDDDSDIGRAGHNM	1957
Qy	1860	RREFESRWEFFYQ	1872
Db	1958	RYEFKKWTDTFK	1970
RESULT 12			
ADP54420			
ID	ADP54420	standard; protein; 1972 AA.	
XX	AC	ADP54420;	
XX	DT	18-NOV-2004 (first entry)	
XX	XX	Human PRO protein sequence SEQ ID NO:396.	
DE	DE	XX	
KW	KW	human; PRO; immune related disease; inflammatory immune response;	
KW	KW	immune response stimulation; antiallergic; antianaemic; antiarthritic;	
KW	KW	antisthmatic; antidiabetic; antiinflammatory; antipsoriatic;	
KW	KW	antirheumatic; antithyroid; CNS; dermatological; gastrointestinal;	
KW	KW	haemostatic; hepatotropic; immunostimulant; immunosuppressive; muscular;	
KW	KW	neurotropic; neuroprotective; osteopathic; respiratory; vasotropic;	
XX	KW	virucide; gene therapy.	
XX	XX		
XX	XX	Homo sapiens.	
XX	XX	WO2004039956-A2.	
XX	XX	13-MAY-2004.	
XX	XX	28-OCT-2003; 2003WO-US034381.	
XX	XX	29-OCT-2002; 2002US-0422472P.	
XX	XX	(GETH) GENENTECH INC.	
XX	PA	Aggarwal S, Clark H, Gurney AL, Schoenfeld J, Williams PM;	
PI	PI	Wood WI, Wu TD;	
XX	XX	WPI; 2004-376182/35.	
XX	XX	N-PSDB; ADP54419.	
XX	XX	New PRO polynucleotides and polypeptides, useful in diagnosing	
PT	PT	and treating an immune related disease, e.g. systemic lupus	
PT	PT	erythematosus, rheumatoid arthritis, diabetes mellitus or asthma and in	
PT	PT	stimulating an immune response.	
XX	XX	Claim 1; SEQ ID NO 396; 3009pp; English.	
XX	XX	The present invention describes an isolated PRO nucleic acid (1). Also	
CC	CC	described: (1) a vector comprising (1); (2) a host cell comprising the	
CC	CC	vector of (1); (3) a process for producing a PRO polypeptides; (4) an	
CC	CC	isolated PRO polypeptide; (5) a chimeric molecule comprising the	
CC	CC	polypeptide of (4) fused to a heterologous amino acid sequence; (6) an	
CC	CC	antibody which specifically binds to a polypeptide of (4); (7) a	
CC	CC	composition of matter comprising a polypeptide of (4), an agonist or	
CC	CC	antagonist of the polypeptide or an antibody that binds to the	
CC	CC	polypeptide in combination with a carrier; (8) an article of manufacture	
CC	CC	comprising a container, a label on the container and a composition of	
CC	CC	matter of (7); (9) a method of treating an immune related disease in a	
CC	CC	mammal; (10) a method for determining the presence of a PRO polypeptide	
CC	CC	in a sample suspected of having the polypeptide; (11) a method of	
CC	CC	diagnosing an immune related disease or an inflammatory immune response	
CC	CC	in mammal; (12) a method of identifying a compound that inhibits or	

CC mimics the activity of or expression of a gene encoding a PRO polypeptide
 CC ; and (13) a method of stimulating the immune response in a mammal. The
 CC PRO sequences have antiallergic, antianemic, antiarthritic,
 CC antiaschemic, antidiabetic, antiinflammatory, antiporiatic,
 CC antirheumatic, antichyroid, CNS, dermatological, gastrointestinal,
 CC haemostatic, hepatocytic, immunosuppressant, immunosuppressive, muscular,
 CC nephrotropic, neuroprotective, osteopathic, respiratory, vasotropic and
 CC virucide activities, and can be used in gene therapy. The nucleic acid
 CC (I) and the encoded polypeptides, compositions, kits and methods are
 CC useful in diagnosing and treating an immune related disease and in
 CC stimulating an immune response. The present sequence represents a human
 CC PRO protein from the present invention.

xx
 SQ Sequence 1972 AA;

Query Match 25.3%; Score 2482.5; DB 8; Length 1972;
 Best Local Similarity 32.2%; Pred No. 3.7e-157;
 Matches 680; Conservative 329; Mismatches 676; Indels 426; Gaps 68;

QY 1 MEMEANEANDHFNFTGLPPAPAAAGLKPS-----SGEGLYTNKSPMFPQ 48
 DB 43 MESSNSDSD---SGTSSDTSSEGISSSDDLEDEEDQSTIESEDDSDSESAQ 98

QY 49 GKSLGDNVNGSLTVSTTTGILNSAPHSSSTSLHHP-SVAYDCLWNTSQPSANPG 107
 DB 99 HKS-NNQVLLHG---ISDPKADG--QKATEKAQEKRIHQPLPLAPE-----SQTHSFQ-- 145

QY 108 SNLKDPPLLSQ-----FSGGVPLNGI-----LGGSRPSS----- 138
 DB 146 SOQKQPVLSQOLPIFQSSQAKESVKNKHTSVIOSTGLVSNKPLSLVNOAKETYNKL 205

QY 139 --PSHNTNLRAGSQKFWANGTHSPMGLNFDSDLPDQ-----NFEVCSGI-- 186
 DB 206 IVPSPDV-LKAGNKNTSESSILTSLSKREYKQAFPSQLKKQESSKSLKVTAAASN 264

QY 187 -----HPDEAAEKMTSVVANGTGLVCSLEEXQBELKMGVNGSVPSVESLHQ 237
 DB 265 PRATSSSPAHPKQTLNHNPNFLNAL-----LGNHP-----NG---VIQSVIQ 307

QY 238 EVSLVPDPTVSCLDPPSHLPDQEDTILSDLSLEPNSLAPEVPVSGLYGIDDTLM- 296
 DB 308 EAPL-----ALTTKTMOSKINEN--IAAASSTPFPSSPVNLTSGRRTPGNQTPVP 357

QY 297 -----GAEDKPLDSDPVISALDCPSLNATAFSILLADDSTISIPASPTSPVL 347
 DB 358 SASPTLHSGQKEKAVSNVNPVT-----QHSHPAKSLVEQFRCTDSIPSSKDSN 412

QY 348 GESVLQDNSFDLNNGSDAQEMETQSDFFPSPLOPA-----PDQSTIQLHPAT 398
 DB 413 EDEEDDEDEDEDEDDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSD 472

QY 399 SPVSPPTSPVSLVSPVSPVSPVSPVSPVSPVSPVSPVSPVSPVSPVSPVSPVSP 458
 DB 473 SMKLKTT-----SVKSPMS----- 489

QY 459 LTASVTPSKASVTPSPAPAFPTASPAKNDVVSFLTTADVEITGEGLTAS--GSDVMRR 517
 DB 490 LTGHTSPNLHITAKAPGA--PAALCSSESQSPALGTSSST-----LTSPPHSQTSKR 541

QY 518 RIATPEEVRLPLQHWRRREVRIKKGSHRWQGTWYTGCGKRMKOPPEVIKYLNSLVHS 577
 DB 542 RVTDERELRIPLQYQWQRETRIRNFGRLQGEVAYAPCGKKLQYQPEVIKYLNSRNGIMD 601

QY 578 VREHFSRPMPCVDFFERTDTPGLQVQLSAEIPSRIOAITGKGRPRNTEKATK 637
 DB 602 ISRDNFSFAKTRVDFEARDGPQEMQWCLKEEDVPIPRAMEGRGRRPNPDRQAR 661

QY 638 EVPKVRGRGRPPVKITLNLKNTNRPLKLEAQE-----TLNEEDKAKIAK 685
 DB 662 EBSRMRRRGRPNVGNAEFLDNAOKLLRKLQAEIARQAAQIKLLRKLQKQEQARVAK 721

QY 686 SKKMR-----QKVQGECLTTIQOQARNKQKQETSLKH-----KEAKKKSAREK 733

DB 722 EAKQQAQMAAEKRRKQKQEKIKMKQKQEKIKRIQOIRMEKELRAQQIIEAKKKKKEEAA 781
 QY 734 GK--TKQKLEK-----VKREKKEKVKM-----KEKEEVTAKACKAD 771
 DB 782 AKLLEAEKRIKEKEMRRQAVLLKHQERRRRQHMMLKAMEARKAEKELKQEKRE 841
 QY 772 KTLATQRLERQKQ 831
 DB 842 KRLNKRKLEQRLEMAKELKPNEDMCLADQKPLPELPRIFGLVLSGSGTFSCLMVV 901
 QY 832 EFLSHFGKVLGDFPAKQVPSGLVQEGLLCGDSISLGEVQDILLVLLKALHDPGPPSCY 891
 DB 902 QLRNFGKVLGDFVNIIDVNLVLSVQEGLLNIGDSMGEVQDILLVLLSAVCDPGLITGVK 961

QY 892 SLKILGEKVSIEPLTRDNVSEILRCFLMAYGVXPALCDRLRTQPFQAPPOKQAAVLAPP 951
 DB 962 AKTALGEHLLNVGNRDNVSEILQIFMEAHCGQTELTESLTKAFAQHTPAKASVLAFL 1021

QY 952 VHELNGSTLIINEIDKLTLESMSYRKNWIVVEGRRLKTVLAKTGRSE-----VEMGRP 1007
 DB 1022 INELACSKSVSEIDKNIDYMSNLRDKWVVEGKRLRIIHAKTGKRTSGGIDLGE 1081

QY 1008 EECGLG-----RRSSRIME-----ETSGWEEEEEESIAAVPGRGR--RDGEV 1049
 DB 1082 QHPLGTPTPGKRRKRGSDYDDDDDDSDGDEDEDEDEDEDEDEDEDEDEDEDEDE 1141

QY 1050 DATASSPELBRQLEKSLKQLFRKLLHSSQMLRAVSLGQDRYRRYVWVLYLAGIFV 1109
 DB 1142 GDQAASVEELEKQLEKSLKQSQYRKLFDASHLSRSMVFGDRYRRYVWVLYPRCGGIFV 1201

QY 1110 EGTEGNLVPVRIKKTDSLKVAHAASLNPAFLSMKMLAGSNNTTASSPARSR-PLKT 1168
 DB 1202 EGMESEGLEE-IAKERELKKAESVQIKEMF-----ETSGDSLNCSTNDHCEQEDLKE 1256

QY 1169 KPG---FWQ-PRHFKSPVRCQDSEQOQALQPEALHVPAPQPOQLQLOLQSHKGFLEQE 1224
 DB 1257 KDNTLNLFQKPGFSKLSKLE-----VAKMPPESEVMTF-----KPNAGAN 1298

QY 1225 GSPLSLGQS--QSAFLSWLSQTSQSHSLSSSVL--TPDSSPGKL-DPAPSOPP 1278
 DB 1299 CCTLISYQNSGKSLGSVQS-----TATQSNVEKADSNLNTGSSGKGFVSLPND-- 1350

QY 1279 EEPDEAESPPDLOAFWFNISAQWPCN-----AAPTPLAVSEDDQPTSPQQ 1326
 DB 1351 ----QLLKTLEKRNQWFSLLPRTPCDDTSLTHADMSTASLVTSPQSPSPSPSPTEAP 1405

QY 1327 LASSKPMRPAANP--CSPVQF--SSTPLAGLAKRAGDPG-----EMPQSTP 1372
 DB 1406 LGSS--AQNPVGLNPFALSPLOKVGVSMMGL---QFCGWPTGVVTSNIPFTLSVPSLGS 1460

QY 1373 GLGQPKRGR-----PPSKF---FKOMEQRYLTQTA-----QPVPPMCS 1410
 DB 1461 GLGSEGNISFLTSNVASSKSESPVQNEKATSQAPAAVEKVPDVPSPKPIEEMQF 1520

QY 1411 GWWIIPDEMLDAMLKALHPRGIREKALHKLHNRHDFLOEVLCRPSADPIFEPPQLPAF 1470
 DB 1521 GWRRIIDPEDLALLKVLHLRGIREKALQKIQKHLDVITQACLKNKQDAVIELNENEN 1580

QY 1471 Q---EGMSWSPKETTETDLAVLOWVEELQSRVIMSDLOIRGWTCPSPDSTREDLAYCE 1527
 DB 1581 QVTRDIVENWVSEEQAMEMDLSVLQVEDLERRVASASLQVKGMCMCPESPASREDLVYPE 1640

QY 1528 HLSDS---QSDITWRGPGREGI-APQKTTNPLDILAVMLAALBNVKKRYRLREPLWPT 1582
 DB 1641 HKSFTLCKEHGDEFTGEDSSAHLEKSNPLDIAVTRLADLERNIERR----- 1691

QY 1583 HEVLEKALLSTPNGAPEGTTTISYEITPRIRIWRQTLQRCRSAHVCCLGLHLSRSTA 1642
 DB 1692 -----FEEDTAPGLRVWRALSEARSAAQVALCICQLQKSLA 1728

QY 1643 WEKSNVKNVTCULVRKGDNEFLLICDGDRCRGCHYIHRPKMEAVPEGWFCVCLAQ--- 1699
 DB 1729 WEKSIMKYVCQICRKGDNELLLDGDGCKGCHYIHRPKITTIIDGDMFCPACIAGSG 1788

Qy	1700	-----QVRGEFTQKPGFKRGQK-----RKSGYSLNFSSEGGRRRRVLLK 1739	CC	preventing or treating metabolic disorders such as obesity or diabetes.
Db	1789	QTLKIKLHVGGKKTNE---SKGKKVLTGDTEDDSASTSSSLKRGNKDLQKKM--- 1842	CC	The polynucleotides are useful as probes or primers. The complex is particularly useful for identifying selected interacting domains (SID (RTM)) for screening drugs that modulate the protein interaction, thus exhibiting the therapeutic effect. The present sequence represents a SID (prey) protein of the invention
Qy	1740	GRESPAAGRYSEERLSKRRRLSMRHHSDLTFCIEILMEMESHDAAPFXEPVNPRL 1799	CC	
Db	1843	-EENTSIINLSKQESFTSVKKPK----RDSKDIALCSMILTEMETHEDAWPFLPVNLKL 1897	CC	
Qy	1800	VSGYRRIIKNPDSTMRERLLRGYTSSEFAADALLVFNQCFTFNEDDSEVGKAGHIM 1859	CC	
Db	1898	VPGYKVKVTKKMPDFSTIREKLSSGQYPNLETALDVLVFNFCFTFNEDDSDIGRAGHNM 1957	CC	
Qy	1860	RRFFESRWEFFQY 1872	CC	
Db	1958	RKIFEKKWTDTFK 1970	CC	
RESULT 13				
ID	ABU70931	standard; protein; 280 AA.	Qy	Query Match 14.5%; Score 1420; DB 6; Length 280;
XX	ABU70931		Db	Best Local Similarity 99.3%; Pred. No. 6.4e-87;
AC	ABU70931		Qy	Matches 278; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
DT	10-JUN-2003	(first entry)	Db	
DE	Human adipocyte	Selected Interacting domain, SID, #562.	Qy	727 SXAEKEGKTKQEKLEKVKREKKEKVKMKEKEVTKAKPACKADKTLATORRLEERQK 786
XX	Human; prey; adipocyte; SID; selected interacting domain; anorectic;		Db	1 SKAEKEGKTKQEKLEKVKREKKEKVKMKEKEVTKAKPACKADKTLATORRLEERQK 60
KW	antidiabetic; protein-protein interaction; diabetes;		Qy	787 QMILEEMKKPTEDMCLTDHQPLPDFSRVPGTLTLPFGAFSDCLTIVEFLHSGKVLGPDPA 846
KX	yeast 2-hybrid assay; metabolic disorder; obesity.		Db	61 QMILEEMKKPTEDMCLTDHQPLPDFSRVPGTLTLPFGAFSDCLTIVEFLHSGKVLGPDPA 120
OS	Homo sapiens.		Qy	847 KDVPESLGVLOEGLLCQGSLSGEVQDLLVRLKKAALHDPGFPSCQSLKILGEKVSEIPLT 906
XX			Db	121 KDVPESLGVLOEGLLCQGSLSGEVQDLLVRLKKAALHDPGFPSCQSLKILGEKVSEIPLT 180
XX			Qy	907 RDNVSEILRCFLMAYGVKPCALCDRLRTQPFQAOQPPQKAAVLAPFVHELNGSTLIINEID 966
XX			Db	181 RDNVSEILRCFLMAYGVKPCALCDRLRTQPFQAOQPPQKAAVLAPFVHELNGSTLIINEID 240
XX			Qy	967 KTLSSMSYRNKNWIVEGRLRLKTVLAKRTGRSEVEMGR 1006
XX			Db	241 KTLSSMSYRNKNWIVEGRLRLKTVLAKRTGRSEVEMGR 280
RESULT 14				
ID	AAB93073	standard; protein; 708 AA.	XX	Human protein sequence SEQ ID NO:11898.
XX	AAB93073		XX	Human; primer; detection; diagnosis; antisense therapy; gene therapy.
AC	AAB93073		XX	Homo sapiens.
DT	26-JUN-2001	(first entry)	XX	EP1074617-A2.
DE	Human protein sequence	SEQ ID NO:11898.	XX	07-FEB-2001.
XX			XX	28-JUL-2000; 2000EP-00116126.
XX			XX	29-JUL-1999; 99JP-00248036.
XX			XX	27-AUG-1999; 99JP-00300253.
XX			XX	11-JAN-2000; 2000JP-00118776.
XX			XX	02-MAY-2000; 2000JP-00183767.
XX			XX	09-JUN-2000; 2000JP-00241899.
XX			XX	(HELI-) HELIX RES INST.
XX			XX	Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
XX			XX	Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX			XX	WPI; 2001-318749/34.
XX			XX	Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.
XX			XX	Claim 8; SEQ ID NO 11898; 2537pp + Sequence Listing; English.


```
QY 1070 QLFFRKKLLHSSQMLRAVSLGQDRYRRYVWLPYLAGIFVEGTGNLVPVEVIKKTDSL 1129
Db 121 QSOYRRKLFDAHSLSRVSFQGDYRRYRWILPQCGGIFVEGMSGEGLER-IAKERKL 179
QY 1130 KVAHAASLNLPALFSMKMELAGSNNTTASSPARASR-PLKTRPG----FMQ-PRHFKSPVRG 1184
Db 180 KKAESVQIKKEMF-----ETSGDSLNCSTNDHCEQKEDLKEKDNNTNLFQKPGSFKLSKL 235
QY 1185 QDSEQPAQLQPEAQLHVPAPQPOLQLQSHKGFLQEGSPLSLQGS-QHDLS--QSA 1241
Db 236 LE-----VAKMPESEVMT-----KPNAGANGCTLSYQNSGKHSLSGSVQS- 276
QY 1242 FLSWLSOTQSHSSLLSSVL--TPDSSPGKL-DPAPSQPPEPEPDEAESPDLOAFWN 1298
Db 277 -----TATQSNVEKADSNLNTGSSGPKFYSPLPND-----QLLTKLTEKNRQWFS 324
QY 1299 ISAQMPCN-----AAPTPLAVSEDOPTSPQQLASSKPMNRPSAANP--CSP 1344
Db 325 LLPRTPCDDTSLTHADMSTASLVTPSQSPPSKSPSPTPAPLGSS--AQNPGVGNPFALSP 382
QY 1345 VQF-SSTPLAGLAPKRAGDP-----GEMPOSPTGLGQPKRGR-----PPS 1385
Db 383 LQVKGVSMMGL---QFCGMPGVVTSNIPFTSSVPSLSGLSGLSEGNNSFLTNSVASS 439
QY 1386 KF---FKQMEORYLTQJTA-----QVPPPEMCSGWWIPDPPEMLDAMLKALHP 1430
Db 440 KSESPVPQNEKATSAQPAAVEAKVPDPSPKPIPEEMQFGWRIIDPEDLKALLKVLHL 499
QY 1431 RGIREKALHKLHNRDFOEVLCLRPSADPIFEPRQLPAFQ---EGIMSWSPKKTETD 1487
Db 500 RGIREKALQKQIQKHLDVITQACLKNKDVAIIELNENEENQVTRDIVENWSVEQAMBD 559
QY 1488 LAVLQWVELQORVIMSLQIRGWTCPSPDSTREDLAYCEHLSDS-----QSDITWRGPR 1543
Db 560 LSVLQVQEDLERRVASASLQVKGWMCPEPASERDLVFEHKSFTKLCKEHDSFTGEDE 619
QY 1544 EGL-APQKTTNPLDLAVWRLAALQONKRYLREPLWPTHEVLEKALLSTPNCAPEGT 1602
Db 620 SSAHALERKSNPLDIAVTRDLADLERNIER----- 650
QY 1603 TTEISYEITPRIRWQTLQRCRSAAHVCLCLGHLERSIAWEKSNKVTCLVCRKGNDNE 1662
Db 651 ---IETDAPGLRVWRRALSARSAAQVALCIQQLQKSIAMEKSIKMYCQICRKGNDNE 707
QY 1663 FLLLCDCGDRGCHIYCHRPKMEAVPEGDWFTCVCLAAQVVEGFTQKPGFPKRGOKRSGY 1722
Db 708 LLLLCDCGDKGCHTYCHRPKITTPDGDWFCFACIA-----KASGQ 748
QY 1723 SLNFSE--GDGRRRRVLLKGRESAAGPRYSEERLSPSKRRRLNRNHSPL 1772
Db 749 TLKIKLHVKGKNTNESKKGRKVLTGTDEDEDSASTSS-----SLKRGNKDL 796

RESULT 16
ADL30971
ID ADL30971 standard; protein; 796 AA.
XX AC ADL30971;
XX DT 20-MAY-2004 (first entry)
XX DE Human protein encoded by a full length cDNA clone SeqID 3004.
XX KW human; medicine; signal transduction; glycoprotein; transcription;
XX KW oligo-capping method.
XX OS Homo sapiens.
XX FN EPI396543-A2.
XX PD 10-MAR-2004.
XX
```

07-JUL-2000; 2003EP-00025638.

08-JUL-1999; 99JP-00194486.

11-JAN-2000; 2000JP-00118774.

02-MAY-2000; 2000JP-00183865.

07-JUL-2000; 2000EP-00114089.

(REAS-) RES ASSOC BIOTECHNOLOGY.

Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
WPI; 2004-204755/20.

N-PSDB; ADL30970.

New oligonucleotide primers (830 cDNAs) useful for synthesizing full
length human cDNAs.

Example 1; SEQ ID NO 3004; 1340pp; English.

This invention relates to a novel primers useful for synthesising full
length cDNA molecules that encode human proteins. Specifically, it refers
to secretory or membrane proteins that are potential therapeutic agents/
target molecules in the field of medicine, and in particular genes
encoding proteins that are associated with signal transduction,
glycoproteins and transcription. The present invention describes a method
for efficiently cloning a full length human cDNA from both the 5' and 3'
ends using the oligo-capping method. This polypeptide sequence is a full
length human protein of the invention.

Sequence 796 AA;

Query Match 11.3%; Score 1108.5; DB 8; Length 796;
Best Local Similarity 33.4%; Pred. No. 2.9e-63;
Matches 298; Conservative 124; Mismatches 283; Indels 187; Gaps 33;

QY 972 MSSYRKWKVTEGRRLKTVLAKTRGRSE----VEMGRPEECIG-----RRSSSRIM 1020
Db 1 MSNLRDKWVVEGKRLKRIIHAKTGRDTSGGIDGEEQHPLGTTTPGKRKRKKGDS 60
QY 1021 E-----ETSGMEEEEEESIAAVPGRGR--RDGEVDATASSIPELERQIEKLSKR 1069
Db 61 DYDDDDDDSDQDQDEDEDEDEKDKGKTDCEDDEGDQAAASVEELEKQIEKLSKQ 120
QY 1070 QLFFRKKLLHSSQMLRAVSLGQDRYRRYVWLPYLAGIFVEGTGNLVPVEVIKKTDSL 1129
Db 121 QSOYRRKLFDAHSLSRVSFQGDYRRYRWILPQCGGIFVEGMSGEGLER-IAKERKL 179
QY 1130 KVAHAASLNLPALFSMKMELAGSNNTTASSPARASR-PLKTRPG---FMQ-PRHFKSPVRG 1184
Db 180 KKAESVQIKKEMF-----ETSGDSLNCSTNDHCEQKEDLKEKDNNTNLFQKPGSFKLSKL 235
QY 1185 QDSEQPAQLQPEAQLHVPAPQPOLQLQSHKGFLQEGSPLSLQGS-QHDLS--QSA 1241
Db 236 LE-----VAKMPESEVMT-----KPNAGANGCTLSYQNSGKHSLSGSVQS- 276
QY 1242 FLSWLSOTQSHSSLLSSVL--TPDSSPGKL-DPAPSQPPEPEPDEAESPDLOAFWN 1298
Db 277 -----TATQSNVEKADSNLNTGSSGPKFYSPLPND-----QLLTKLTEKNRQWFS 324
QY 1299 ISAQMPCN-----AAPTPLAVSEDOPTSPQQLASSKPMNRPSAANP--CSP 1344
Db 325 LLPRTPCDDTSLTHADMSTASLVTPSQSPPSKSPSPTPAPLGSS--AQNPGVGNPFALSP 382
QY 1345 VQF-SSTPLAGLAPKRAGDP-----GEMPOSPTGLGQPKRGR-----PPS 1385
Db 383 LQVKGVSMMGL---QFCGMPGVVTSNIPFTSSVPSLSGLSGLSEGNNSFLTNSVASS 439
QY 1386 KF---FKQMEORYLTQJTA-----QVPPPEMCSGWWIPDPPEMLDAMLKALHP 1430
Db 440 KSESPVPQNEKATSAQPAAVEAKVPDPSPKPIPEEMQFGWRIIDPEDLKALLKVLHL 499
QY 1431 RGIREKALHKLHNRDFOEVLCLRPSADPIFEPRQLPAFQ---EGIMSWSPKKTETD 1487


```
QY 1038 VGRGRGRDGEVDATASSIPELQ-----IEKLSKROLFFRKLLHSSQ 1082
D 1038 VGRGRGRDGEVDATASSIPELQ-----IEKLSKROLFFRKLLHSSQ 1082
Db 859 KGRGRGKG-----QNGKFTRQEQINCVTRELITADEEBALQEHQORKEKLEKIQ 912
QY 1083 MLRAVS-----LCQDRYRRYVWLPYLAGIFVEG-----TEGNLVPEEVIKKETDSLKVA 1133
D 1083 MLRAVS-----LCQDRYRRYVWLPYLAGIFVEG-----TEGNLVPEEVIKKETDSLKVA 1133
Db 913 SAIACTNIFPLGRDRMYRYWIFPSIPLFTBEDYSGULTEMLP----- 957
QY 1134 HASLNPALFSMKWELAGSNNTASSPARARSRLKTKPGFMQPRHFKSPVRGQDSQPQQA 1193
D 1134 HASLNPALFSMKWELAGSNNTASSPARARSRLKTKPGFMQPRHFKSPVRGQDSQPQQA 1193
Db 958 -----RPSFQNNVQSD----- 970
QY 1194 LOPEAQLHVPAPQPOLQLQSHKGFLEQEGSLGQSHDLSQSFLSWLSQTSQHS 1253
D 1194 LOPEAQLHVPAPQPOLQLQSHKGFLEQEGSLGQSHDLSQSFLSWLSQTSQHS 1253
Db 971 -----FQV-----STKIGE 979
QY 1254 SILSSSVLTDPSSPKLPAPSPQPEEPDEAESPDLOAFWFNISAQPCNAAPTPL 1313
D 1254 SILSSSVLTDPSSPKLPAPSPQPEEPDEAESPDLOAFWFNISAQPCNAAPTPL 1313
Db 980 PLMSESTNIDQPR-----DHSVQLP----- 1001
QY 1314 AVSEDPQTPSPQOLASSKPMNRSAANPCSPVQFSSTPLAGLAPKRAGDPGEMPQSP 1373
D 1314 AVSEDPQTPSPQOLASSKPMNRSAANPCSPVQFSSTPLAGLAPKRAGDPGEMPQSP 1373
Db 1002 -----RPVHKPR----- 1009
QY 1374 LQPPRRGRPRSPKFKQMEQRYLTQTAQPVPEMCSGMMWITPDPMLDAMLKALHPRGI 1433
D 1374 LQPPRRGRPRSPKFKQMEQRYLTQTAQPVPEMCSGMMWITPDPMLDAMLKALHPRGI 1433
Db 1010 -----WCFYSSCEQLDQIEALNSRGH 1031
QY 1434 REKALHKLHNRDQFQVCLRPSPADPFI-----EPROLPAQEGIMS--WSPKPKTYE- 1485
D 1434 REKALHKLHNRDQFQVCLRPSPADPFI-----EPROLPAQEGIMS--WSPKPKTYE- 1485
Db 1032 RESALKETLLQEKSIQALQARFSEKTHFSKQPKQDSKPTYSRGRSSNAYDPSQMAEK 1091
QY 1486 TDLAVLQWVELEQRTVM-----SDLQIRGWTCPSPD-----STRE 1521
D 1486 TDLAVLQWVELEQRTVM-----SDLQIRGWTCPSPD-----STRE 1521
Db 1092 QLELRDLFDIEDRIYQGTGAIKVTDRHTRSALESRYELLSEENKENGIIKTNYE 1151
QY 1522 DLAYCEHLSDSQEDITWGPGRGELAPQKTNP-----LDLAVMLAALFQNVKR 1572
D 1522 DLAYCEHLSDSQEDITWGPGRGELAPQKTNP-----LDLAVMLAALFQNVKR 1572
Db 1152 DVEENEIDQTKVIVKDRLLGKTETPTSVINASTPQSVSVVHYLAWALFQIEQIER 1211
QY 1573 RYLREPL-----WPTHEVLEKALLSTNGAPEGTTTETISVITPRIRWQTLQRCRSA 1627
D 1573 RYLREPL-----WPTHEVLEKALLSTNGAPEGTTTETISVITPRIRWQTLQRCRSA 1627
Db 1212 RFLKAPLDASDGRSYKTVLDR-----WRESLLSSASL 1244
QY 1628 AHVCLICHLERSIAWEKSVNVTCLVCRKGNDFELLCDGCDRGCHYCHRPQWEAVP 1687
D 1628 AHVCLICHLERSIAWEKSVNVTCLVCRKGNDFELLCDGCDRGCHYCHRPQWEAVP 1687
Db 1245 SOVFLHLSTLDRSVIWSKILNARCKICRKGDAENMYLDCGCDRGHTYCVRPKLKTPV 1304
QY 1688 EGDWFCVCLAQ-----VEGEFTQKPGFKRGOKRKGYSILN 1725
D 1688 EGDWFCVCLAQ-----VEGEFTQKPGFKRGOKRKGYSILN 1725
Db 1305 EGDWFCPCRPQRCERLSFRQPSLSEDEVEDSMGGEDVEDGDEEGQSEEEYVE 1364
QY 1726 FSEGD-----GR-----RRVLLKGR-----ESPAAGPRYSE----- 1752
D 1726 FSEGD-----GR-----RRVLLKGR-----ESPAAGPRYSE----- 1752
Db 1365 QEDDSQEEEVSLPKRGQVRLPVKTRGLSSFSRGGQQEQGPRYPSRQSTPTKT 1424
QY 1753 -----ERLSPSKRR----- 1762
D 1753 -----ERLSPSKRR----- 1762
Db 1425 VSSKTRSLRKINSAPPTETSLRIASRSTRHSHGFLQADVVELLSPRKRGRKRSANN 1484
QY 1763 -----LSMRNHS----- 1770
D 1763 -----LSMRNHS----- 1770
Db 1485 TPENSPNPNFNVATKSEGRSVNIASKLSQESKRRKCRKQSPSPVTLGRSS 1544
QY 1771 -----DLTFCEIILMEMESHDAAPFXEPVNPRLVSGYRRIKPNMDFSTMRLRLRG 1823
D 1771 -----DLTFCEIILMEMESHDAAPFXEPVNPRLVSGYRRIKPNMDFSTMRLRLRG 1823
Db 1545 GRQGVHLSAFEQLVVLVRHDDSWFLKLVSKIQVPDYDIKKPIALNIIRKVNKC 1604
QY 1824 GYTSSEFAADALLVDFNQCOTFNEDDSVEVGKAGHIMRFF 1863
D 1824 GYTSSEFAADALLVDFNQCOTFNEDDSVEVGKAGHIMRFF 1863
Db 1605 EYKLASEFIDITELMFNSCFEYNPNRNTSEAKAGTRLQAFF 1644
```

```
RESULT 21
AAB54103
ID AAB54103 standard; protein; 141 AA.
XX
AC AAB54103;
XX
DT 09-MAR-2001 (first entry)
XX
DE Human pancreatic cancer antigen protein sequence SEQ ID NO:555.
XX
KW Human; pancreas; pancreatic cancer; pancreatic cancer antigen; detection;
diagnosis; identification; cytostatic; neuroprotective; nootropic;
immunomodulatory; relaxant; contraceptive; gynaecological;
antiinflammatory; cardiant; gene therapy; chromosome mapping;
linkage analysis; tissue identification; tissue typing; forensic; neural;
immune system; muscular; reproductive; gastrointestinal; pulmonary;
cardiovascular; renal; proliferative.
XX
OS Homo sapiens.
XX
WO200055320-A1.
XX
PD 21-SEP-2000.
XX
PF 08-MAR-2000; 2000WO-US005989.
XX
PR 12-MAR-1999; 99US-0124270P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM;
XX
WPI; 2000-579444/54.
XX
N-PSDB; AAC98868.
XX
New nucleic acid that is a pancreatic cancer antigen for preventing,
treating, or ameliorating a medical condition, particular pancreatic
cancer, or for use in assays for diagnosing a pathological condition.
XX
Claim 11; Page 994; 1379pp; English.
XX
AAC98773 to AAC99231 encode the human pancreatic cancer associated
proteins, called pancreatic cancer antigens, given in AAB54008 to
AAB54466. The human pancreatic cancer antigens have cytostatic,
neuroprotective, nootropic, immunomodulatory, relaxant, contraceptive,
gynaecological, cardiant and antiinflammatory activities, and can be used
in gene therapy. The polynucleotide and proteins can be used for
preventing, treating, or ameliorating a medical condition or in assays
for diagnosing a pathological condition or a susceptibility to one in a
subject. Binding partners to the proteins and the activity of the
proteins can be identified. The pancreatic cancer antigens can be used to
detect, treat or prevent pancreatic disorders, especially cancer.
XX
Agonists and antagonists to the antigens can be screened for. The
pancreatic cancer antigen polynucleotides can be used to design nucleic
acid hybridization probes that can be used in chromosome mapping, linkage
analysis, tissue identification and/or typing and a variety of forensic
and diagnostic methods. The proteins can be used to generate antibodies
which are used to purify, detect and target the polypeptides, including
both in vivo and in vitro diagnostic and therapeutic methods. The
proteins can be used to treat or prevent neural, immune system, muscular,
reproductive, gastrointestinal, pulmonary, cardiovascular, renal or
proliferative disorders. AAC99232 to AAC99240 and AAC94467 represent
sequences used in the exemplification of the present invention
XX
Sequence 141 AA;
```

```
Query Match 5.9%; Score 575; DB 3; Length 141;
Best Local Similarity 81.8%; Pred. No. 1.7e-30;
Matches 108; Conservative 9; Mismatches 11; Indels 4; Gaps 1;
QY 492 LETTADVEITGEGITAGSGDVMRRRIATPEEVLPLQHGWRREVRKKGSHRWQGTW 551
|||||
```

Db 4 LETTADVBEITGEGUTASGGDNVMMRRRIATPEEVRLPLQHGWRREVRIKSGSHRWQGETW 63

QY 552 YGPGCGKMKOPPEVIKYLNRNLVHSVREHFSPRMPVGDFFPERDTPEG-----LOWV 607

Db 64 YGPGCGKMKOPPEVIKYLNRNVHSVREHFSPRMPVGDFFPERKHARGADPKVYA 123

QY 608 QLSAEEIPSRIQ 619

Db 124 FVPEELVDKLQ 135

RESULT 22

ID AAB63263

XX AAB63263 standard; protein; 244 AA.

AC AAB63263;

XX 26-MAR-2001 (first entry)

DT Human breast cancer associated antigen protein sequence SEQ ID NO:625.

DE Human; breast cancer; gastric cancer; prostate cancer; diagnosis;

KW cancer associated antigen; cytostatic; cancer vaccine.

XX Homo sapiens.

OS

XX WO200073801-A2.

PN

PD 07-DEC-2000.

XX 26-MAY-2000; 2000WO-US014749.

PF

XX 28-MAY-1999; 99US-0136526P.

PR 10-SEP-1999; 99US-0153454P.

XX

XX (LUDW-) LUDWIG INST CANCER RES.

PA

XX Obata Y;

PI

XX WPI; 2001-025274/03.

DR

XX Nucleic acids encoding breast, gastric and prostate cancer associated

PT antigen precursors, useful for diagnosing and treating a condition

PT characterized by expression of an abnormal amount of a protein, e.g.

PT cancer.

XX

PS Example 1; Page 488-489; 799pp; English.

XX

XX AAF22422 to AAF22626, AAF22627 to AAF22773 and AAF22774 to AAF23014

CC represent nucleotide sequences encoding human breast, gastric and

CC prostate cancer associated antigen precursors (CAAP) respectively.

CC AAB63232 to AAB63467, AAB63468 to AAB63721 and AAB63722 to AAB63970

CC represent human breast, gastric and prostate CAAP protein sequence

CC respectively. CAAPs have cytostatic activity and can be used in the

CC production of cancer vaccines. The human CAAP proteins, peptides, nucleic

CC acids or anti-CAAP antibodies are useful for diagnosing and treating a

CC condition characterised by expression of an abnormal amount of a protein,

XX e.g. cancer

XX

XX Sequence 244 AA;

QY Query Match 5.7%; Score 561; DB 4; Length 244;

Best Local Similarity 41.8%; Pred. No. 3.5e-29;

Matches 109; Conservative 37; Mismatches 65; Indels 50; Gaps 4;

QY 1625 RSAHVCLGLHLERSIAWEKSVNKTCLVCRKGDNDFLLICDCCRGCHYCHRPKWE 1684

Db 3 RSAQVALCICQOLQKSIWEKSIKMYCQICRKGDNBEILLICDCCDKGCHYCHRPKIT 62

QY 1695 AVPEGDWFCTVCLAQOVGEFTQKPGFKRGQRKSGYSLNFSF--GGRRRRVLLKGRE 1742

Db 63 TIPDGDWFCAPACIA-----KASGQTLKIKLHVKGKKTNSKGGK 103

QY 1743 SPAAGPRYSBERLSPS-----KRRRLS-----MRNHSDLT 1773

Db 104 VTLTGDTDEDSASTSSSLKRGKNDLKKRMEENTSINLSKQESFTSVKKPKRDDSKDLA 163

QY 1774 FCEILMESHDAAPXEPVNPRLVSGYPRRIIKNPMDFSTMERLLRGGYTSEEFAA 1833

Db 164 LCSMILTEMETHEDAWPFLPVNLKLVPGYKKVIKKPMDFSTIREKLSGGQYPNLETAL 223

QY 1834 DALLVFNCFQTFNEDDSEVGK 1854

Db 224 DVRLVFDTVKHLMEDDDSDIGR 244

RESULT 23

ID ADQ20910

XX ADQ20910 standard; protein; 1483 AA.

AC ADQ20910;

XX 26-AUG-2004 (first entry)

DT Human soft tissue sarcoma-upregulated protein - SEQ ID 3730.

DE Human soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human.

KW soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human.

XX Homo sapiens.

OS

XX WO2004048938-A2.

PN

PD 10-JUN-2004.

XX 26-NOV-2003; 2003WO-US038193.

PF

XX 26-NOV-2002; 2002US-0429739P.

PR

XX (PROT-) PROTEIN DESIGN LABS INC.

PA

XX Aziz N, Ginsburg WM, Zlotnik A;

PI

XX WPI; 2004-441208/41.

DR

XX Early detection of soft tissue sarcoma comprises determining expression

PT of a gene in a first soft tissue sample and a normal soft tissue sample

PT and comparing the gene expression, also useful in treating soft tissue

PT sarcoma.

XX

PS Example 2; SEQ ID NO 3730; 210pp; English.

XX

XX The invention relates to a novel method for detecting soft tissue sarcoma

CC which comprises obtaining a first soft tissue sample from an individual

CC and a normal soft tissue sample from the same or different individual,

CC determining the expression of a gene in both samples and comparing the

CC expression of the gene in both soft tissue samples, where a higher level

CC of protein expression in the first soft tissue sample indicates the

CC presence of soft tissue sarcoma. The method of the invention has

CC cytostatic applications and may be useful for detecting soft tissue

CC sarcoma, possibly via gene therapy or vaccine production. The nucleic

CC acid sequences may be useful in diagnostic and screening applications.

CC The current sequence is that of a human soft tissue sarcoma-upregulated

CC protein of the invention. The current sequence is not shown within the

CC specification per se but was submitted in CD format by the inventor.

XX

XX Sequence 1483 AA;

QY Query Match 5.4%; Score 528.5; DB 8; Length 1483;

Best Local Similarity 19.5%; Pred. No. 8.3e-26;

Matches 265; Conservative 179; Mismatches 439; Indels 479; Gaps 42;

QY 619 QAITCKGRPR---NTEKATKEVPKVRGRGPRPKVKITEL-----LNKTDNRPLKKL 669

Db 415 QKSTGNSKSPKGLKTPKTKMKQMTLLDMAGTQ---KMTAPRNSGGTPTRTSKPHKHL 471

QY 670 EA-----QETLNEEDKAKIAKSKKMRQKQVORGECLTTIQQARNKRKQETKSLKH 720

PT nucleic acid with a test agent and detecting a test agent-biased
PT activity.

XX Example; SEQ ID NO 110; 406pp; English.

XX The present invention describes a method for identifying a candidate p53
CC pathway modulating agent, which comprises: (a) providing an assay system
CC comprising a MP53 (modulator of p53) polypeptide or nucleic acid or its
CC fragment or derivative; (b) contacting the assay system with a test agent
CC under conditions where the system provides a reference activity except in
CC the presence of the test agent; and (c) detecting a test agent-biased
CC activity, where a difference between the test agent-biased activity and
CC the reference activity identifies the test agent as a candidate p53
CC pathway modulating agent. Also described: (1) modulating the p53 pathway
CC of a cell; (2) modulating the p53 pathway in a mammalian cell; and (3)
CC diagnosing a disease in a patient. MP53 has cytostatic activity, and can
CC be used in gene therapy. The method is useful for identifying a candidate
CC p53 pathway modulating agent for preparing a composition for diagnosing
CC or treating e.g., cancer. The present sequence represents a human MP53
CC protein, which is used in the exemplification of the present invention.

XX Sequence 1483 AA;

Query Match 5.3%; Score 525.5; DB 7; Length 1483;

Best Local Similarity 19.5%; Pred. No. 1.3e-25;

Matches 265; Conservative 178; Mismatches 440; Indels 479; Gaps 42;

QY 619 QAITCKGRPR---NTEAKTEKVPKVGRCRPPKVKITEL-----LNKTDNRPLKL 669
DB 415 QKSTGNSKSPKGLTKPTKMKOMTLLQWAKGTQ---KMTAPRNSGGTPTRTSSPKHKL 471
QY 670 EA-----QETLNEEDKAKIAKSKKMKQKQVQGECLTTIQCAQNRKQKQETKSLKH 720
DB 472 PPAALHLIAYKENKDREKRSALSCVTSKTRALLSSEDRAPLPELSRLVQKRYELLEH 531
QY 721 KE--AKKXSABEKKGTKQEKLEKVKREKKEKVKMEKEKEVTKAKPACKADKTLATOR 778
DB 532 KKRWASMEEQREKYLKKREELKKLEKAKER---REKE-----569
QY 779 LLEERQKQMIILEMKKPTEDNCLTHQPLDPFSVPGLT-LPGAFSDCLTIVEFLHSP 837
DB 570 MLERLEKQ-----KRYEDELTYG-KNLPAFLVOTPEGLNTLFGDVAWVFLSCY 620
QY 838 GKVLFDPDAKQVPSGLVQELGLCCQDSIGEVQDILLVRLKALHDPGPPSYCQSLKITLG 897
DB 621 SGLLLPDAQYPTAVS-LMEALSADKGGFLYLNRLVLVLTQLLQDEIAEDYGE----LG 675
QY 898 EKVSIEPLTRONVSEILRCFLMAYG-----VXPALCDRLRTQ 934
DB 676 MKLSIEPLTHSVSELVLCRLRRSDVQEESESGSDTDNKNDSAAFEDNEVQDEFLEKLETS 735
QY 935 PFOAQPPQOKAAVLAPPHVHENGSTLIINEIDKTESMSSYRKKNKWIVEGRLRLK---- 990
DB 736 EFFETSEKQILTALCHRI---LMTYSVDHMETRQQMSAELW--KERLAVLKEEND 789
QY 991 TVLAKRTGRSEVEM-----GRPECLGRRRRSRIMEETSGMEEEESIAAPVCR--- 1042
DB 790 KKAEEKQKKEAKKNGKVGKGLTKDRKKEIVKPEPQVDYTAEDWISAVKSRLLA 849
QY 1043 --GRBDGVATASSIPELERQIEKLSKR-----QLPFRKXLLSHSQMLRAVSLGQDQY 1094
DB 850 IQAKKEREIQEREMKV-KLERQAEERIRKHAARAKAFQEGIAKAKLVMRRTPGTORN 908
QY 1095 RRRYVWL-PYLAGIVFEGTEGHLVPEEVIKKTDSLKVAAHASINPALFPMQWELAGNT 1153
DB 909 HNRVYLFSDVPLGLFIE-----KGWVHDSID-----934
QY 1154 TASSPARARSRLTKTPGFMQPHFKS-PVRGDSQEQPOALQPEAQHLVPAQPOQLQ 1212
DB 935 -----YRFNHCKDHTVSGDEYCFASK-----957
QY 1213 QIQSHKGFLQEGSPSLQSQSHDLQSQAFLSWLSQTSQSHSLSSSVLTPDSSPGKLPD 1272

DB 958 -----KANLGNK-----ASMNTQHGCTATATEVAVETTTPKQG----- 987
QY 1273 APSQPPEPEDEAESSPDLQAFWFNISAQWPCNAAPTPLPLAVSESDQTPSPQQLASSKP 1332
DB 988 -----QNLWFLCDSQ----- 997
QY 1333 MNRPSAANPCSPVQFSSTPLAGLAPKRRAGDPGMPQSPGTLGQPKRRGRPPSKFFKQME 1392
DB 998 ----- 997
QY 1393 QRYLTQLTAQVPVPPCMCGMWIIPDPEDMLKALHPRGIREKALHKLHANK-HRDFLQE 1451
DB 998 -----KELDELLNCLHPQIGRESQKLERLKRKYODIHS 1031
QY 1452 VCLRPSADPIPEPRLQPAFQEGIMSWSPKE-----KTYETDVLAVLWVVELEORVIMSD 1505
DB 1032 IHL--ARKPNLGLKSCDGNQE-LNLFRLSDLIETVATRLQKGLGVVETSEFEARVIS-- 1086
QY 1506 LOIRGWTCPSPDSTREDLAYCEHLSDSQEDITWRGPGREGIAPORKTTNPLDLAVMLAA 1565
DB 1087 -----LEKLKDFGECV-----IA 1099
QY 1566 LEQNVKRYLREPLWP-----THEVLEKALLSTPNGAPEGTTTETSYEIT 1611
DB 1100 LQASVIKKFLQGFVAPKQKRRKLQSEDSAKTEEVEDEEKQWVE-----EAKVA 1146
QY 1612 PRIRIWRQTLQCRSAHVCICLGHLESIKAWKSVNVTCLVCRKNGNDLFLILCDGCD 1671
DB 1147 SALEKWKTAIREAQTFSRMHVLLGWLDCIKWDSANARCKVCRKKGEDDKLILCDECN 1206
QY 1672 RGCHYCHRPKQWAEVPEGDMFTVC-----LAQVQVEGFTQKGPFPK 1713
DB 1207 KAFHLFCLRLPALYVPPDGEWQCPACQAPATARRNSGRNRYTESASEDSEDESEDEEE 1266
QY 1714 RGQRKSGSYLNFSEGDGRRRRLVKGRES---PAA--GPR-----YSEERLSPSK---- 1759
DB 1267 EEEEEEEDYEV---AGLRLRPKRTIRGKHSVIPPAARSGRRPGKPKPHSTRSQPKAPPVD 1323
QY 1760 -----RRRLSNRNHSDLTFCETILMEMESHDAAPFPKVPNPRLVSGYRRIKNP 1810
DB 1324 DAEVDELVLQYKRSRRSQSLQKCEELHLMIVKRFSPWFPREPVRDEAEDYDVIHP 1383
QY 1811 MDFSTMRERLRGGYTSEEFAADALLVDFNCOTFNEDDDSEV 1852
DB 1384 MDFQTQVQKCGSYRSVQEFELTDMKQVFTNAEVNCRGSHV 1425
RESULT 26
AAW81172
ID AAW81172 standard; protein; 1527 AA.
XX
AC AAW81172;
XX
DT 05-MAY-1999 (first entry)
XX
DE Human BAZ1-beta protein #1.
XX
KW Transcriptional regulator; transcription; BAZ1-alpha; bromodomain; BAZ;
KW atypical zinc finger; testis; human; tumour; BAZ1-beta; BAZ2-alpha; drug;
KW BAZ2-beta; treatment; cancer; proliferative disorder; screening.
OS Homo sapiens.
XX
FN W09847920-A1.
XX
PD 29-OCT-1998.
XX
PF 17-APR-1998; 98WO-JP001783.
XX
PR 18-APR-1997; 97JP-00116570.
PR 24-OCT-1997; 97JP-00310027.
XX
PA (CHUS) CHUGAI RES INST MOLECULAR MEDICINE INC.

XX Jones MH;
XX
XX
DR WPI: 1998-583603/49.
DR N-PSDB; AAV68408.
XX
PT Transcriptional regulator gene family containing bromodomain - may be
PT expressed in testis tissue and is useful for treatment of cancer and
PT other proliferative disorders.
XX
XX
PS Claim 1; Page 125-137; 187pp; Japanese.
XX
CC This sequence represents the human BAZ1-beta protein, a member of a
CC family of transcriptional regulator genes containing a bromodomain (BAZ,
CC Bromodomain with Atypical Zinc finger) which are expressed specifically
CC in testis tissue and also in certain tumour lines. Transgenic cells may
CC be used for the preparation of the BAZ1-alpha, BAZ1-beta, BAZ2-alpha and
CC BAZ2-beta proteins. These proteins can be used in the treatment of cancer
CC and other proliferative disorders, and in screening of compounds for
CC their binding ability to the expression products (e.g. for use as drugs
CC by modulation of transcriptional regulation)
XX
SQ Sequence 1527 AA;
Query Match 5.3%; Score 521.5; DB 2; Length 1527;
Best Local Similarity 19.4%; Pred. No. 2.6e-25;
Matches 264; Conservative 179; Mismatches 440; Indels 479; Gaps 42;
QY 619 QAITCKGRPR---NTEKAKTKEVPKVRGRGPRPKVKITEL-----LNKTDNRPLKL 669
DB 415 QKSTGNSPKPKGLTKPTKMKQMTLLDMAGTQ---KMTAPRNSGGTPTRTSSPKHKL 471
QY 670 EA-----QETLNEEDKAKIAKSKKKQKQVORGECLTIOQOARKKQKETSLSKH 720
DB 472 PPAALHLIAYYKENDKREDKSAISCVISKTARLSSDDRARLPPEELSLVQKRYELLEH 531
QY 721 KE--AKKKSXAEEKGTKQEKLEKVKREKKEKVKMEKEEVTAKPACKADKTLATQR 778
DB 532 KRWASMEEQKEVLKKRRELKLEKAKER---REKE----- 569
QY 779 RLEERQKQMLEENMKKPTDMCMCTDHPDPFSRVPGLT-LPGAFSDCLTIVEFLHSF 837
DB 570 MLERLEKQ-----KRYEQDELGT-KNLPAPFLVDTPEGLNFTLFGDVAWVVEFLSCY 620
QY 838 GKVLGFDPAKDVPSLGVLOEGLLCOGDSLGVEODLLVPLLAALHDPFPFYCQSLKILG 897
DB 621 SGLLLPDAQYPIAVS-LMEALSADKGGFLYLNRLVILLOTLQDEIAEYGE----LG 675
QY 898 EKVSIEPLTRDNVSEILCFMAYG-----VXPALCDRLRTQ 934
DB 676 MKLSIEPLTSHVSELVRLCLRRSDVQESSEGSDDTDNKSNAAPEDNEVQDEFLEKLETS 735
QY 935 PFQAQPOOKAAVLAPVHVLNGSTLIINEIDKLTESMSSYRNKNKWIIVEGRRLRK---- 990
DB 736 EFFELTSEKQLILTALCHRI---LMTYSVDHMETRQMSAELW--KERLAVLKEEND 789
QY 991 TVLAKTGRSEVM-----GRPECLGRRRSRMEETSMEEBEESIAAIVGRR--- 1042
DB 790 KKRAEKQKREMEAKNKENGKVGNGLTDRKRIVKFPQVDTEAEDMI SAVKSRLLA 849
QY 1043 --GRRDGEVDATASSIPELERQIEKLSK-----QLFFRKLLHSSOMLRAVSLGDQRY 1094
DB 850 IQAKKERIQREMKV-KLERQAEERIRKHAARAKAFQEGIAKAKLVMTPTIGTDNRN 908
QY 1095 RRRYVVL-PYLAGIFVEGTEGNLVPEEVIKKETDSLKVAHAASLNPAFLFSMKMELAGSNT 1153
DB 909 HNRVWLFSDVEPGLFIE-----KGWVHDSID----- 934
QY 1154 TASSPARARSPLTKTPCFMQPRHFKS-PVRGQDSEQPAQLQPEAQLHVAQPQPLQL 1212
DB 935 -----YRFNHCKDHTVSGDEDYCPRSK----- 957
QY 1213 QLQSHKGFLQEGSPLSQHQHDLQSASFLSWLSQTSQSHSSLLSSSVLTPDSSPGKLDP 1272

DB 958 -----KANLGKN-----ASMTQHGCTATEVAVETTPKQG----- 987
QY 1273 APSQPPEPEDEAESPDLOAFWNFISAQMPCNAAPTPPLAVSEDDQTPSPQQLASSKP 1332
DB 988 -----QNLWFLCDSQ----- 997
QY 1333 MNRPSAANPCSPVQFSSTPLAGLAPKRRAGDPGEMPSQPTGLQPKRRGRPPSKFFKQME 1392
DB 998 ----- 997
QY 1393 QRYLTQLTAAQVPVPEMCGMWWIPDPPEMLDAMLKALHPRGIREKALHKHLNK-HRDFLQOE 1451
DB 998 -----KELDELLNCLHPQGIRESQKLERLEKRYQDIHS 1031
QY 1452 VCLRPSADPTPEPQLPAFQBGIMSWSPKE-----KTYETDLAVLQWVELEORVIMSD 1505
DB 1032 IHL--ARKPNLGLKSCDGNQE-LLNFLRSDLIEVATRLQKGLGYVEETSEFEARVIS-- 1086
QY 1506 LQIRGWTCPSPDSTREDLAYCEHLSDSOEDITWRGPGREGLAQPKRTTNPLDLAVMLAA 1565
DB 1087 -----LEKUKDFGECV-----IA 1099
QY 1566 LEQNVKRYRLREPLW-----THEYVLEKALISTPNGAPEGTTTISYBIT 1611
DB 1100 LOASVIKKFLQGFMAQKQKRKLQSEDSAKTEEVDEEKQV-----EAKVA 1146
QY 1612 PRIRWQTQORCSAAHVCLIGHLERSIAWEKSVNKKVTCLVCRKGDNDDEFLLCDCGD 1671
DB 1147 SALEKWKTAIREAQTSRSMHVLLGMLDACIKWDMSAENARCKVCPKKGEDDKLILCDSCN 1206
QY 1672 RGCHYCHRPKMEAVPEGDMFCTVC-----LAQQVEGFEFTQKPGFPK 1713
DB 1207 KAPHLFCURPALYEVDPGEWQCPACQATARNRSGRNYTEESASEDSEDEEBEEEE 1266
QY 1714 RGQRKSGYSILNFSEGDGRRRRVLLKGRES---PAA--GPR-----YSEERLSPSK---- 1759
DB 1267 EEEEBEEDYEV---AGLRPRKRTIRGKHSVIPPAASGRPGKKPHSTRSQPKAPPVD 1323
QY 1760 -----RRRLSMRNHSDLTFCETIILMEMESHDAAPFKXPVNPRLVSGYRRIKNP 1810
DB 1324 DAEVDELVLQTKRSRRRQSLQCEILKHKIVKFRFSWPPREFVTRDEADYDYDVIHP 1383
QY 1811 MDFSTMERLLRGYTSSEEPFAADALLVFDCNCTFNEDDSEV 1852
DB 1384 MDFQTVQNKSCGYSRVSQVEFLTDKQVFTNAEVNCRGSHV 1425
RESULT 27
AAW81173
ID AAW81173 standard; protein; 1531 AA.
XX
AC AAW81173;
XX
DT 05-MAY-1999 (first entry)
XX
DE Human BAZ1-beta protein #2.
XX
KW Transcriptional regulator; transcription; BAZ1-alpha; bromodomain; BAZ;
KW atypical zinc finger; testis; human; tumour; BAZ1-beta; BAZ2-alpha; drug;
KW BAZ2-beta; treatment; cancer; proliferative disorder; screening.
XX
OS Homo sapiens.
XX
FN W09847920-A1.
XX
PD 29-OCT-1998.
XX
PF 17-APR-1998; 98WO-JP001783.
XX
PR 18-APR-1997; 97JP-00116570.
PR 24-OCT-1997; 97JP-00310027.
XX

(CHUS) CHUGAI RES INST MOLECULAR MEDICINE INC.

Jones MH;

WPI: 1998-583603/49.

N-PSDB; AAV68409.

Transcriptional regulator gene family containing bromodomain - may be expressed in testis tissue and is useful for treatment of cancer and other proliferative disorders.

Claim 1; Page 145-157; 187pp; Japanese.

This sequence represents the human BAZ1-beta protein, a member of a family of transcriptional regulator genes containing a bromodomain (BAZ, Bromodomain with Atypical Zinc Finger) which are expressed specifically in testis tissue and also in certain tumour lines. Transgenic cells may be used for the preparation of the BAZ1-alpha, BAZ1-beta, BAZ2-alpha and BAZ2-beta proteins. These proteins can be used in the treatment of cancer and other proliferative disorders, and in screening of compounds for their binding ability to the expression products (e.g. for use as drugs by modulation of transcriptional regulation)

Sequence 1531 AA;

Query Match 5.2%; Score 512.5; DB 2; Length 1531;
Best Local Similarity 19.2%; Pred. No. 1e-24;
Matches 262; Conservative 179; Mismatches 446; Indels 475; Gaps 41;

619 QAITGKRRPR---NTEKAKTEVKVKGGRGPPKVKITEL-----LNKTDNRPLKLL 669
619 QAITGKRRPR---NTEKAKTEVKVKGGRGPPKVKITEL-----LNKTDNRPLKLL 669
415 QKSTGNSKPKKGLTPKTKMQLMDWAKGTQ---KMTAPRNSGGTPTTSSPKHKL 471
415 QKSTGNSKPKKGLTPKTKMQLMDWAKGTQ---KMTAPRNSGGTPTTSSPKHKL 471
670 EA-----OETNEEDKAKIAKKKMKORQVORGECLTTIQGAARKKQKQETSLKH 720
670 EA-----OETNEEDKAKIAKKKMKORQVORGECLTTIQGAARKKQKQETSLKH 720
472 PPAALHLIAYKENDREKRGALSCVSKTARLLSSDRARLPBELSLVKRYELLEH 531
472 PPAALHLIAYKENDREKRGALSCVSKTARLLSSDRARLPBELSLVKRYELLEH 531
721 KS--ARKKSABKKGKTKQEKLEKVKREKKEKVKMEKEBEVTKAKPACKADKTLAQR 778
721 KS--ARKKSABKKGKTKQEKLEKVKREKKEKVKMEKEBEVTKAKPACKADKTLAQR 778
532 KRWASMSSEQRKEVKKKREELKKLEKAKER---REKE----- 569
532 KRWASMSSEQRKEVKKKREELKKLEKAKER---REKE----- 569
779 RLEERQKQMIILEMKKPTEDMCLTDHQLPDPFSRVPLGT-LPSGAFSDCLTIVFLHSP 837
779 RLEERQKQMIILEMKKPTEDMCLTDHQLPDPFSRVPLGT-LPSGAFSDCLTIVFLHSP 837
570 MLERLEKQ-----KRYEDQELTG-KNLPAPRLVDTPEGLPNTLFGDVAMVFFLSY 620
570 MLERLEKQ-----KRYEDQELTG-KNLPAPRLVDTPEGLPNTLFGDVAMVFFLSY 620
838 GVLGFDPAKQVPSLGVLOEGLLCQDLSLGEVQDILLVRLKALHDPGPPSYCQSLKILG 897
838 GVLGFDPAKQVPSLGVLOEGLLCQDLSLGEVQDILLVRLKALHDPGPPSYCQSLKILG 897
621 SGLLLPDAQYPITAVS-LMEALSADKGGFLYLNRLVILLQTLQTLQDETAEDYGEIG 679
621 SGLLLPDAQYPITAVS-LMEALSADKGGFLYLNRLVILLQTLQTLQDETAEDYGEIG 679
898 EKVSIPITRDNVSEILRCFLMAYG-----VXPALCDRLRTQ 934
898 EKVSIPITRDNVSEILRCFLMAYG-----VXPALCDRLRTQ 934
680 MKLSIPIPLTHSVSELVRLCLRRSDVQEESEGDDTDNKSAAAFEDNEVQDFLEKLETS 739
680 MKLSIPIPLTHSVSELVRLCLRRSDVQEESEGDDTDNKSAAAFEDNEVQDFLEKLETS 739
935 PFQAPQPOQKAAVLAPPVHELNGSTLIINEIDKTLSESSSYKRNKWIVEGRRLRK---- 990
935 PFQAPQPOQKAAVLAPPVHELNGSTLIINEIDKTLSESSSYKRNKWIVEGRRLRK---- 990
740 EFFELTSEBKQILTALCHRI----LMTYSVDHMETROQMSAELW--KERLAVLKEEND 793
740 EFFELTSEBKQILTALCHRI----LMTYSVDHMETROQMSAELW--KERLAVLKEEND 793
991 TVLAKRTGRSEVM-----GRPEELGRRSSRIMEETSGMEEBEESIAAPVGR--- 1042
991 TVLAKRTGRSEVM-----GRPEELGRRSSRIMEETSGMEEBEESIAAPVGR--- 1042
794 KGRABQKQKMEAKNKNKGVENGKGTDRKKRIVKFPQVDVTAEDMISAVKSRLLA 853
794 KGRABQKQKMEAKNKNKGVENGKGTDRKKRIVKFPQVDVTAEDMISAVKSRLLA 853
1043 --GRDGEDVATASSIPELQIEKLSK---QLFPRKLLHSSQMLRAVSLGQDRY 1094
1043 --GRDGEDVATASSIPELQIEKLSK---QLFPRKLLHSSQMLRAVSLGQDRY 1094
854 IQAKEREIQOEREMKV-KLERQAEERINFKHAAAEKAFQEGIAKAKULVMRTPTGDRN 912
854 IQAKEREIQOEREMKV-KLERQAEERINFKHAAAEKAFQEGIAKAKULVMRTPTGDRN 912
1095 RRRYWL-PYLAGI FVEGTEGNLVPEEVIKKETDSLKVAHAASLNPALFSPMOMELAGSNT 1153
1095 RRRYWL-PYLAGI FVEGTEGNLVPEEVIKKETDSLKVAHAASLNPALFSPMOMELAGSNT 1153
913 HNRVWLFSDVPGLEFIE-----KGWVHDSID----- 938
913 HNRVWLFSDVPGLEFIE-----KGWVHDSID----- 938
1154 TASSPARARSRLPKTKPGMQPRHFKS-PVRQDSEQPAQLOPEAQLHVPAPQPOLQL 1212
1154 TASSPARARSRLPKTKPGMQPRHFKS-PVRQDSEQPAQLOPEAQLHVPAPQPOLQL 1212
939 -----YRFNHCKDHTVSGDEYCPRSK----- 961
939 -----YRFNHCKDHTVSGDEYCPRSK----- 961

QY 1213 QLQSHKGFLEQEGSPLSLGQSOHDLSSQAFLSWLSQTSQSHSSLLSSSVLTDPSSSGKLDLP 1272
QY 1213 QLQSHKGFLEQEGSPLSLGQSOHDLSSQAFLSWLSQTSQSHSSLLSSSVLTDPSSSGKLDLP 1272
Db 962 -----KANLGKN-----ASMTQHGATATEVAVETTTPKQG----- 991
Db 962 -----KANLGKN-----ASMTQHGATATEVAVETTTPKQG----- 991
QY 1273 APSQPPEPPEDEAESPDLOAFWFNISAQWPCNAAPTPLAVSEDQTPPSQQGLASSKP 1332
QY 1273 APSQPPEPPEDEAESPDLOAFWFNISAQWPCNAAPTPLAVSEDQTPPSQQGLASSKP 1332
Db 992 -----QNLWPLCDSQ----- 1001
Db 992 -----QNLWPLCDSQ----- 1001
QY 1333 MNRPSAANPCSPVOFSSTPLAGLAPKRRAGDPGEMPOSPTGLGQPKRRRPPSPKFKOME 1392
QY 1333 MNRPSAANPCSPVOFSSTPLAGLAPKRRAGDPGEMPOSPTGLGQPKRRRPPSPKFKOME 1392
Db 1002 ----- 1001
Db 1002 ----- 1001
QY 1393 QRYLTQLTAQVPPPEMCSGWWIPDPPEMLKALHPRGIREKALHKHLNK-HRDFLOE 1451
QY 1393 QRYLTQLTAQVPPPEMCSGWWIPDPPEMLKALHPRGIREKALHKHLNK-HRDFLOE 1451
Db 1002 -----KELDELLNCHLPQIGRESOLKERLEKRYQDIHS 1035
Db 1002 -----KELDELLNCHLPQIGRESOLKERLEKRYQDIHS 1035
QY 1452 VCLRPSADPIPEPQPLPAFOFGINSWSPKE-----KTYETDLAVLQVVEELEQEVINDS 1505
QY 1452 VCLRPSADPIPEPQPLPAFOFGINSWSPKE-----KTYETDLAVLQVVEELEQEVINDS 1505
Db 1036 IHL--ARKPNLGLKSCDQNB-LNLFRLSDLIEVATRIQKGGGLGYVEETSEFEARVIS-- 1090
Db 1036 IHL--ARKPNLGLKSCDQNB-LNLFRLSDLIEVATRIQKGGGLGYVEETSEFEARVIS-- 1090
QY 1506 LQIRGWTCPSPDSTREDLAYCEHLSDSQEDITWRGPGREGLAPQRKTTNPLDLAVMLAA 1565
QY 1506 LQIRGWTCPSPDSTREDLAYCEHLSDSQEDITWRGPGREGLAPQRKTTNPLDLAVMLAA 1565
Db 1091 -----LEKLDKFGECV-----IA 1103
Db 1091 -----LEKLDKFGECV-----IA 1103
QY 1566 LEQNVKRRYLREPLWP-----THEVVLEKALLSTPNGAPEGTTTETSYEIT 1611
QY 1566 LEQNVKRRYLREPLWP-----THEVVLEKALLSTPNGAPEGTTTETSYEIT 1611
Db 1104 LQASVIKKFLQGFMAKQKRRKLOSEDSAKTEEVDEEKKMVE-----EAKVA 1150
Db 1104 LQASVIKKFLQGFMAKQKRRKLOSEDSAKTEEVDEEKKMVE-----EAKVA 1150
QY 1612 PRIRWRTLQCRSAAHVCLCLGHLERSIAWEKSVNKTCLVCRKGDNDDELLACDGC 1671
QY 1612 PRIRWRTLQCRSAAHVCLCLGHLERSIAWEKSVNKTCLVCRKGDNDDELLACDGC 1671
Db 1151 SALEKWKTAIRAEQOTFSRMHVLLGMLDACIKWDSAEANARCKVCPKGGEDDKLIJCDCN 1210
Db 1151 SALEKWKTAIRAEQOTFSRMHVLLGMLDACIKWDSAEANARCKVCPKGGEDDKLIJCDCN 1210
QY 1672 RGCHYCHRPKMEAVPEGDWFCVTC-----LAQQVEGEFTQKPGFPK 1713
QY 1672 RGCHYCHRPKMEAVPEGDWFCVTC-----LAQQVEGEFTQKPGFPK 1713
Db 1211 KAFHLFLCLRPALYEVPDGEWQCPACQATARNRSGRNYTESASESDESDDESEEE 1270
Db 1211 KAFHLFLCLRPALYEVPDGEWQCPACQATARNRSGRNYTESASESDESDDESEEE 1270
QY 1714 RGQKRSYSLNFBEGDGRRRRLVLLKGRES---PAA--GPR-----YSEERLSPSK---- 1759
QY 1714 RGQKRSYSLNFBEGDGRRRRLVLLKGRES---PAA--GPR-----YSEERLSPSK---- 1759
Db 1271 EEEBEEEDYEV---AGLRPRKRTIRGKHSVTPPAARSGRRPGKPHSTRSOPKAPPVD 1327
Db 1271 EEEBEEEDYEV---AGLRPRKRTIRGKHSVTPPAARSGRRPGKPHSTRSOPKAPPVD 1327
QY 1760 -----RRRLSMRNHSDLTFCETIILMEMESHDAAPFXYEPVNPRLVSGYRIIKNP 1810
QY 1760 -----RRRLSMRNHSDLTFCETIILMEMESHDAAPFXYEPVNPRLVSGYRIIKNP 1810
Db 1328 DAVIDELVLTQKRSRRQSLQKCEILHKIVKRYFSWPRFPVTRDEAEDYDVITHP 1387
Db 1328 DAVIDELVLTQKRSRRQSLQKCEILHKIVKRYFSWPRFPVTRDEAEDYDVITHP 1387
QY 1811 MDFSTMRLRLRGYTSSEFAADALLVFDNCQTFENEDDSEV 1852
QY 1811 MDFSTMRLRLRGYTSSEFAADALLVFDNCQTFENEDDSEV 1852
Db 1388 MDFQTVQNKCGSYRSVQEFELTDMKQVFTNAEVNCRGSHV 1429
Db 1388 MDFQTVQNKCGSYRSVQEFELTDMKQVFTNAEVNCRGSHV 1429

RESULT 28

AAO03463

ID AAO03463 standard; protein; 108 AA.

XX AAO03463;

AC AAO03463;

DT 06-NOV-2001 (first entry)

XX Human polypeptide SEQ ID NO 17355.

XX Human polypeptide SEQ ID NO 17355.

XX Human polypeptide SEQ ID NO 17355.

XX Human polypeptide SEQ ID NO 17355.

XX Human polypeptide SEQ ID NO 17355.

XX Human polypeptide SEQ ID NO 17355.

XX Human polypeptide SEQ ID NO 17355.

XX Human polypeptide SEQ ID NO 17355.

XX Human polypeptide SEQ ID NO 17355.

XX Human polypeptide SEQ ID NO 17355.

XX Human polypeptide SEQ ID NO 17355.

XX Human polypeptide SEQ ID NO 17355.

XX Human polypeptide SEQ ID NO 17355.

XX Human polypeptide SEQ ID NO 17355.

XX Human polypeptide SEQ ID NO 17355.

XX Human polypeptide SEQ ID NO 17355.

XX Human polypeptide SEQ ID NO 17355.

XX Human polypeptide SEQ ID NO 17355.

XX Human polypeptide SEQ ID NO 17355.

XX Human polypeptide SEQ ID NO 17355.

XX Human polypeptide SEQ ID NO 17355.

XX Human polypeptide SEQ ID NO 17355.

XX Human polypeptide SEQ ID NO 17355.

Db 1240 -----IKLKNVTDKKNRESEKKGQR--TSTFQINGKDNKPKIYLKGECLKEISES 1290
Qy 308 PVISALDCPSLNATAFSILLADDSQTSISIFASPTSPVLGSEVLQD-----NSPD---L 359
Db 1291 RVVSGNVEPKVNNIN--KIIPENDIKSLTVKESAIRPFINGDIVINEDFERNSSSTKSHL 1348
Qy 360 NNGSDAE---QREME-----TQSSDFFPPLTOPA---PDQSSSTIQLHP---ATSPAVSPTTS 407
Db 1349 LSSSDAEGNYRDSLETSLSTKESDSTQITTPSASCPSNSVNOVEDMEIETSEVKVKVTS 1408
Qy 408 PAVSLVSPAASPEIS-----PEVCPAASTVV---SPAVFSVVS 443
Db 1409 PITSEENLNSDFIDENGLPIKNENVNGESKRKTVITEVTTMTSTVATESKTVIKVEK 1468
Qy 444 PASSAVLPAVSLVPLTASVTSFKASPTVSPAAPPTASPAKOVSSFLETTADVEEITG 503
Db 1469 GDKQTVSTENCASKSTVTTTITVKLUSTSTGSGVDIIISVKEOSKTVTTTITVDSLTT 1528
Qy 504 EGLTAGSGDVMRRRIATPEEVRLPLQHGWRREVRIRKKGSHRWQGETWYVPCGKRMKQF 563
Db 1529 TGGTLVTSMTV--SKEYSTRDKVL---MKFSRPKXTRSGT-----AL 1566
Qy 564 PEVIKYLGRNLVHSV-----RRE--HFSF-----SPRMPVGD 593
Db 1567 PSYRKFTVKSTKKSIFVLNDLKKLARKGGIREVYFENYNAKPALDIPVPSRPRTFGI 1626
Qy 594 FFEER-DTPEGLQ-----WVOLSAAEIPSRIOAITGKRGRPNTEKAKTKEVPKVK 643
Db 1627 TWYRLQTVKSLAGVSLMRLWLASLRDWDMAKVP-----PCGGSTRTETSET----- 1675
Qy 644 RGRGPPPKVKITELNKTDNRP-----LKKLEAQETLNEEDKAKIAKSKKXMRQKVQ 695
Db 1676 -----BITTEILIKRVDGYPGIRFEYCIKIIICPIGVETPKETTPQKGLRSSAL 1728
Qy 696 RGECLTTIQQARNRKQETSLKHKEAKKXSAEKEKGTQKQELKVKREKKEKVKM 755
Db 1729 R-----PKRPETPKQTGPVITETWABEELELWEIRAFARVEKEKAQAVEQ 1775
Qy 756 KEKEVTKAKPACKADKTATORRLEERQKOOILBEMKPTEDMCLTDHQLPDPFSRVP 815
Db 1776 QAKKRL-----EQQKPTVIATSTTSPTSTST----- 1803
Qy 816 GLTLPAGAFSDCLTIVEFLHSFGVLGPDPAKDPVSLGVLOEGLLCCQDSLSGEVDLLVR 875
Db 1804 -----ISPAQKVMVAPI--SGSVTTGTKM----- 1825
Qy 876 LLKAALHPDPGPSYQCSKILGEKVSEIPLTRDNVSEILRFLMAYGVXPCALDRLRTP 935
Db 1826 VLTTKVGSPATVTFQONKNFHQTATVWKQGSN--SGVVQVQKVLGIIPSGSTGTSQ-QT 1883
Qy 936 FOAQPPOQKAAVLAFVPHVHELNGSTLI INEIDKTLSSMSSYRKWKWIVEGLRR-----LK 990
Db 1884 FTSFOP-RTATVTTIRPNTSGSGT-----TSNSQVITGPQIRPGMTVIR 1926
Qy 991 TVLAKRT--GRS-----EVENGRPECLGRRSSRIMEETSGMEEEESIGIAAVPGR-- 1042
Db 1927 TPLQSTLIGKAIIRFVMVQAPQVM-----TQIIRGQPVSTAVSAPNTVSTSPGQS 1981
Qy 1043 ---GRRDGEVDATASSIPELERQIEKLSKROLFFRKLLHSSQMLRAVSLGQDRYRRYW 1099
Db 1982 LTSATSTSNIOSSASQPPRQOGQVKLTMAQLTQLTQGHGNGQLTTVIQQGGOTTGQLQ 2041
Qy 1100 VLPYLAGIFVEGTEGN-----LVPEVVIK-----ETDSLKVAHAASL 1137
Db 2042 LIP--QGVTVLPGCQQLMQAMPNGTVQRELFTEPLATTATATTTTITVTTAAGTGEQ 2099
Qy 1138 NPALFSMKMELAGSNTTASSPARARS--RPLTKPGFMOPRHFKSP--VRGQDSEOPQAOQL 1195
Db 2100 RQSKLSQPMQVHODKTL--PPAQSSVGPAPAKQPTAQPSARPQOTQPSAPQPEVOTQ 2157
Qy 1196 PEAQL-----HVPAQPOFQLOLOLQSHKGFLEBGSPLSLQSGQSHDLS--QSAFLSWLSQ 1248
Db 2158 PEVQTQTTVSSHVPEAQP---THAQSSKQPVAASQPSQNSVQGSFVRVQSP-----SQ 2209

RESULT 11

ADF42724
ID ADF42724 standard; protein; 2781 AA.
XX ADF42724;
XX AC ADF42724;
XX DT 26-FEB-2004 (first entry)
XX Human BPTF amino acid sequence SEQ ID NO:30.
DE diabetic; pre-diabetic; Type 2 diabetes; antidiabetic; gene therapy;
XX diabetes; insulin resistance; metabolic disease; human.
XX Homo sapiens.
OS
XX WO2003102163-A2.
FN
XX 11-DEC-2003.
PD
XX 04-JUN-2003; 2003WO-US017825.
PF
XX 04-JUN-2002; 2002US-0385857P.
PR 04-JUN-2002; 2002US-0386013P.

PR 04-JUN-2002; 2002US-0386074P.
 PR 04-JUN-2002; 2002US-0386107P.
 PR 05-JUN-2002; 2002US-0386314P.
 PR 05-JUN-2002; 2002US-0386326P.
 PR 05-JUN-2002; 2002US-0386332P.
 PR 05-JUN-2002; 2002US-0386481P.
 PR 05-JUN-2002; 2002US-0386512P.
 PR 05-JUN-2002; 2002US-0386513P.
 PR 05-JUN-2002; 2002US-0386558P.
 PR 05-JUN-2002; 2002US-0386600P.
 PR 05-JUN-2002; 2002US-0386615P.
 PR 05-JUN-2002; 2002US-0386654P.
 PR 06-JUN-2002; 2002US-0386838P.
 PR 06-JUN-2002; 2002US-0386861P.
 PR 06-JUN-2002; 2002US-0386944P.
 PR 06-JUN-2002; 2002US-0386955P.
 PR 06-JUN-2002; 2002US-0387017P.
 PR 06-JUN-2002; 2002US-0387026P.
 PR 06-JUN-2002; 2002US-0387039P.
 PR 20-JUN-2002; 2002US-0386865P.
 XX (META-) METABOLEX INC.
 XX
 XX Allan B, Gregoire F, Lavan B, Moodie S, Waters S, Wong C;
 XX
 XX WPI; 2004-053469/05.
 DR N-PSDB; ADF42723.
 DR
 XX
 PT Identifying an agent for treating diabetic or pre-diabetic individuals
 PT comprises contacting an agent with a polypeptide, e.g., human ceramidase,
 PT and selecting an agent that modulates the expression or activity of the
 PT polypeptide.
 XX
 PS Claim 1; SEQ ID NO 30; 209pp; English.
 XX
 CC The present invention describes a method for identifying an agent for
 CC treating a diabetic or pre-diabetic individual. The method comprises
 CC contacting an agent to a mixture comprising a polypeptide encoded by a
 CC nucleic acid that hybridizes under stringent conditions to a nucleic acid
 CC encoding any of the 23 fully defined amino acid sequences given in the
 CC specification, and selecting an agent that modulates the expression or
 CC activity of the polypeptide. Also described: (1) a method of treating a
 CC diabetic or pre-diabetic animal, comprising administering to the animal a
 CC therapeutic amount of an agent identified by the method described above;
 CC (2) a method of introducing an expression cassette into a cell,
 CC comprising introducing into the cell an expression cassette comprising a
 CC promoter operably linked to a polynucleotide encoding a polypeptide,
 CC where the polynucleotide hybridizes under stringent conditions to a
 CC nucleic acid encoding the above amino acid sequences; and (3) a method of
 CC diagnosing an individual who has Type 2 diabetes or is pre-diabetic,
 CC comprising detecting in a sample from the individual the level of a
 CC polypeptide or the level of the above-mentioned polynucleotide encoding
 CC the polypeptide, where a modulated level of the polypeptide or
 CC polynucleotide in the sample compared to a level of the polypeptide or
 CC polynucleotide in either a lean individual or a previous sample from the
 CC individual indicates that the individual is diabetic or pre-diabetic. The
 CC method is useful in diagnosing and treating diabetes, insulin resistance
 CC or related metabolic diseases in human subjects. The method may also be
 CC used in identifying agents for treating diabetic or pre-diabetic
 CC individuals. The present sequence is used in the exemplification of the
 CC present invention.
 XX
 SQ Sequence 2781 AA;
 Query Match 4.1%; Score 406.5; DB 8; Length 2781;
 Best Local Similarity 18.3%; Pred. No. 3.5e-17;
 Matches 370; Conservative 298; Mismatches 775; Indels 579; Gaps 82;
 QY 79 SSSHLHHPVSAYDCLWYNSQVPSANPGSNLKDPPLLSOFSGGQVPLNGILGGRQPS 138
 DB 1108 SSSKALH-----SSVKS-----TNDRDATPLSR-----AMDPEGLGCDSENS 1148
 QY 139 PSHTNLRAGSKFWANGTHSPMGLNFDQSLEYLDFPDQNFECVSGIHPDEAEKEMTS 198

Db 1149 TLENSSTVSIQ-----DSSEEDMIQNSNESISQFRTRE-ODV-----EVLEPLKCE 1196
 QY 199 VVAENGTLVCSLEEXQPELKMCGYNSVPSVESLHQEYSVLVPDPTVSCLDPPSHLP 258
 Db 1197 LVSGESTG-----NCEORLP-VKGTENGKPKSQOKLEE-----RPVKCSQD----- 1239
 QY 259 DOLEDTPILSEDSLEPNSLAPRPVSGGLYGIDTELMGAEDKLP-----LXDS 307
 Db 1240 -----IKLNTTDDKKNENRESEKQOR--TSTFQINGKDNKPKIYLKGECLKEISES 1290
 QY 308 PVISALOCPSLNNATAPSLADDSQTSISIFASPTSPVLGESVLOD-----NSPD-----L 359
 Db 1291 RVSGNVEPKYNNIN--KIIPENDIKSLTVKESAIRPPINGDVMEDFNERNSSTKSHL 1348
 QY 360 NNGSDAE--OEEME--TQSDPPPSLTOPA--PDOSSTIQLHP-----ATSPAVSPPTS 407
 Db 1349 LSSSDAEGNYRDSLETLPSTKESDSTQTTTSCSPESNSYNQVEDMEIETSEVKKVTS 1408
 QY 408 PAVSLVSPAAASPEIS-----PEVCPAASTVV--SPAVFSVVS 443
 Db 1409 PITSEESNLNSDFIDENGLPINKNENVNGESKRTVITEVTMTSTVATESKTIVKEK 1468
 QY 444 PASSAVLPVLSLEVPLTASVTSPKASPTSPAAAPPTASPAKDVSSFLETTADVEETG 503
 Db 1469 GDKOTVVSSTENCAKSTVTTTTTTLSTSTGSGVDIISVKBSQSKTVTTVTVDLSLT 1528
 QY 504 EGLTASGSDVMRRRIATPEEVRLPLQHGMRREVRIRKKGSHRQOETWYVPCGRMKQF 563
 Db 1529 TGGTLVTSMTV-SKEYSTRDKVL--MKFSRPKTRSGT-----AL 1566
 QY 564 PEVIKYLSRNLVHSV-----RRE--HPSF-----SPRMPVGD 593
 Db 1567 PSYKFKVTSTKSGIFVLNDDLLKLARKGGIRPEVPYFNNAKPAIDWVPSPRPTFGI 1626
 QY 594 FFER-DTPEGLQ-----WVQLSABEETPSRIQAITGKRGPRNTEKAKTREVPKVK 643
 Db 1627 TWRYRLQTVKSLAGVSLMLRLMLASLRWDMAAKVP-----PGGGSTRTETSET----- 1675
 QY 644 RGRGRPPKVKITELLNKTDRNP-----LKLKLEAQETLNEEDKAKIAKKKKRQKVQ 695
 Db 1676 -----EITTEILIKRDVCPYGRFECYRIRKICPGIVPETPKETTPQKGLRSSAL 1728
 QY 696 RGECLTTIQQOARNRKOETKSLKHKEAKKXSAEKEGKTQKQKLEKVKREKEKVKM 755
 Db 1729 R-----PKRPETPKQTGPVITETWAEELWEIRAFARVEKEKAQAVEQ 1775
 QY 756 KEKEEVTYKAPACKADKTLATQRLEROKQOMILEEMKKTEDMCLTDHQPLDPFSRVP 815
 Db 1776 QAKKRL-----EQQKPTVIATSTTSTSTST----- 1803
 QY 816 GLTLPAGAFSCLTIVEFLHSFGKVLGFDPAKDPVSLGVLOEGLLCOQSDSLGEVDLLVR 875
 Db 1804 -----ISPAQKVMVAPI--SGSVTTGTM----- 1825
 QY 876 LLKALHDGPFPSYCSQSLILGEKVEISELPLTRDNVSEILRCFLMAYGVXPALCDRLRTP 935
 Db 1826 VLTTKVSGPATVTFQONKNFHQTATWVKQGSN-SGVVQVQOKVLGIIPSTSTGTSQ-QT 1883
 QY 936 FOADPPQOKAAVLAPPVHELNGSTLIINEIDKTLSEMSSYRKNKVIWEGRLRR-----LK 990
 Db 1884 FTSQP-RTATVTIRPNTSGSGGT-----TSNSQVITGPQIRPGMTVIR 1926
 QY 991 TVLAKRT-GRS-----EVENGRPEECIGRRRRSSRIMEETSGMBEEEEESIAAIPGRR- 1042
 Db 1927 TPLQOSTLGKALITPVMVQGAFCQVM-----TQIRGQPVSTAVSAPNTVSTSPGKS 1981
 QY 1043 ---GRRDEVDATASSIPELEROIEKLSKROLFFRKLLHSSQMLRAVSLGQDRYRRYW 1099
 Db 1982 LTSATSTNIOSSASQPPRQOGQVKLTMAQLTQTOGHGNGQGLTVVIQOGQGTGLOQ 2041
 QY 1100 VLPYLAGIFVSGTEGN-----LVPPEVIKK-----FTDSLKVAAHASL 1137

2042	LIP--QGVTVLPGBGQQLMQAAMPNGVTQRFLFTPLATTATTASTTTTTTSTTAAGTGEQ	2099
1138	NPALF5MKMELAGSNTTASSPARARS-RPLKTKPGFMQRHFKSP-VRGQDSBPQOALQ	1195
2100	RQSKLSFQMQVHQDKTL--PPAQSSVGPAKAQPTAQPSARQPQTQPSPAQPEVQTQ	2157
1196	BEAQL-----HVPAAQPOPQLQLOLQSHKGFLEQEGSPLSLGQSQHDLS-QSASFWSLSQ	1248
2158	PEVQTQTTVSHVPSAQP---THAQSSKQPOVAAQSQPOSNVQSQSPVRVQSP-----SQ	2209
1249	TQSHSLLSSSVLTPDS-SPG---KLPDAPSQP-PEPEPDEAESSPDLQAFWNISAO	1303
2210	TRIPSS-----TPSQLSFGQSQVQTTTISQIPQP-----HSLQI	2246
1304	PCNAAPPTPLAVSEDBQTPSPQOLASSKPMWRPSAANPCSP-VQFSGTPLAGLAPKRRAG	1362
2247	PSQOPQ-----SQPVQSSQTLLSSQOTLNQVSSPSRPFQI-----	2286
1363	DPGEMQSPGLGQPKERGPPSKFFKQMEQRVLTQLTAQVPVPMCSGHWIIPDPEMLD	1422
2287	---QOPQ-POVIAVPQIQOQ--QVLSQIQSQVVAQIQAOQ-----SG---VPQIKLQ	2331
1423	AMLKALHPRGIREKALHKLHKNHRDFLQEVCLRPSADPIPEPQLPAFOGIMSWSPK	1482
2332	LPIQIQSSAVQTHIQINVTVQAASVOE-----QLQRVOQ-LRDQOQKK	2376
1483	TYETDL-----AVLOWVEELEQRVIMSDLIQIRQWTCPSDPSTREDIAYC-----	1526
2377	QQQIEIKREHTLQASNOSEIIOQVNMKNHAVIEHLKKQKSMTPAEREENQRMIVCNQVM	2436
1527	---EHLSDSQEDITWRGPGREGLAPORKTNPDLDLAVMLAALQO-----NVKRRYL---	1575
2437	KYILDKIDKEKQAAKKRKEESVEQKSKONATKLSALLFKKEQURAEILKRRALLDK	2496
1576	-----REPLWPTHEVVLEKALLS-----TPNGAPEGTTTE	1605
2497	DLQIEVQEELKROLKIKKEKDLMLQATAVAAPCPVTPVLPAPPAPPSPPPPGVQH	2556
1606	ISVEITPRIIRWQTLQOR-----CRSAHVCLCLGHLESTAWKSVNKTCLVCRKG	1658
2557	TGLLSTPTLPVASKRKRKEEKSSSKKKKKMISTTSKE-----TKQTKLYCICKTPY	2611
1659	DNDEFLLLCDCGRGCHIYC---HRPKMEAVPBGDMFCTVCLAQVQEGEFTOKGFPFKRG	1715
2612	DESKFYIGCDRCQWYHGRCVGILQSEAEILDE--YVCPQC--QSTEDATV-----	2659
1716	QKRKSGYSLNFSEGDGRRRRVLLKGRSIPAAGPRYSEERLSPSKRRLLSRNHHSDLTFC	1775
2660	-----LTPUTEK-----DYEGL	2671
1776	EILMEMESHDAAMPFXEPVNPRLVSGYRRRIIKNPMDFSMRERLLRGYTSSEFAADA	1835
2672	KRVLSRLOAHKMWPFLEPVDNDPADDYGVIKEPMDLATWEERVQRYEYKLTFFVADM	2731
1836	LLVPDNCQTFNEDDSEVGKAGHIMRRPFESRWEFVQKQAN	1877
2732	TKIFDNCRYNPSDPSFYQCAEVLESFFVOKLKGFKASRSHN	2773

RESULT 32
AD000984

ADO00984 ID ADO00984 standard: protein: 2781 AA.

XX
AC ADO00984:XX
DT 01-JUL-2004 (first entry)

DE Human homologue of Fruit fly AD-related protein CG17135. XX

Human; Alzheimer's disease; Gamma secretase; Pan gene; P-element; EP;
 APP-L-SV; Amyloid precursor-like protein; APP;
 suppressor of hairless transcription factor; Su(H);
 v16 activation domain; dementia; memory loss; language deterioration;

KW	impaired visuospatial skill.
XX	
XX	Homo sapiens.
OS	
XX	
XX	US2004067535-A1.
PN	
XX	
XX	08-APR-2004.
PD	
XX	
XX	03-OCT-2002; 2002US-00263929.
XX	
PF	
XX	
XX	03-OCT-2002; 2002US-00263929.
PR	
XX	
XX	(LIFE-) LIFE SCI DEV CORP.
PA	
XX	
XX	Kim J, Galant R;
PI	
XX	
XX	WPI; 2004-355296/33.
DR	
DR	N-PSDE; ADO00882.
XX	
XX	
PT	Identifying compound by exposing cell that expresses gene having
PT	enhancing or suppression effect on APP1-SV phenotype to agent,
PT	identifying modulation of Alzheimer's disease (AD), regulation of gene or
PT	protein expression with AD.
XX	
XX	
PS	Claim 18; SEQ ID NO 122; 185pp; English.

The invention relates to identifying a compound comprising exposing cell expressing gene 1 having enhancing or suppression effect on an APPL-SV phenotype (a transgenic fruit fly expressing the Amyloid precursor-like protein, APP, as a fusion protein with the suppressor of hairless transcription factor , Su(H) and VPL6 activation domain. The fusion protein is cleaved by gamma secretase (encoded by the Pen gene) to release the Su(H-VPL6, which affects wing vein development. Genes affecting Pen expression/activity were screened by crossing the APP-SV line with an EP-p element insertion library, and the DNA recovered from the appropriate EP strain and sequenced) chosen from ADO00963-ADO00964, being the identified fruit fly genes affecting APP processing and their mammalian homologues, identifying modulation of Alzheimer's disease (AD) symptom, regulation of biological pathway, gene expression or protein function associated with AD relative to cell in absence of agent. Also included are regulating AD (involves providing a subject with AD or symptoms of AD and an agent that changes the expression of a gene described above or changes the activity of a polypeptide having a sequence chosen from ADO00965-ADO01066, and treating the subject with the agent) and a composition (comprising a nucleic acid encoding a polypeptide detailed above or an expression vector comprising the nucleic acid or a host cell comprising the expression vector or an antisense oligonucleotide that hybridises under stringent conditions to the nucleic acid or polypeptide or an antibody that specifically binds to the polypeptide). The method is useful for identifying compounds modulating symptom of Alzheimer's disease (AD), regulation of biological pathway associated with AD, or regulation of gene expression or protein function of gene or protein associated with AD. The nucleic acids and proteins are useful in drug screening and useful in screening and treating the subject having increased susceptibility to AD or symptoms of AD such as dementia, memory loss, language deterioration and impaired visuospatial skills. The present sequence is a human homologue of a fruit fly protein from a gene identified as having an effect on the APP-SV phenotype.

XX Sequence 2781 AA:

Query Match 4.1%; Score 406.5; DB 8; Length 2781;
Best Local Similarity 18.3%; Pred. No. 3.5e-17;
Matches 370; Conservative 298; Mismatches 775; Indels 579;

Qy	79	SSSTSHLHPBSVAYDCLWNYSQVPSNPNGLKDPPLLSQFSGGQVPLNGILGSRQPS	138
Db	1108	SSSKSALH-----SSVPKS---TNRDATPLSR-----AMDFFGLGACDSESN	1148
Qy	139	PSHNTNLRAQSQKFANGTHSPMGLNPFDSQELVDSFPDQNFBEVCSGIHPDEAAEKMT	198
Db	1149	TLENSPTVIGIO---DSSREDIMVONSIESIGEOFTRE-ODV-----EVLPLKCE	1196

Db 1745 IIKRDRVGPYIRSEYCIKRIICPIGVPETPKETPTQKGLRSSLAR----- 1792
Qy 709 NKROETKSLXKHAKEKXSAKEKXKTKQEKLEKVKREKKEKVEVTKAKPAC 768
Db 1793 --PKRPETPKQTPGVIIETWVAEELELWEIRAFARVERKEKAQAVEQAKKRL----- 1844
Qy 769 KADKTLAQRLEERQKQOMLEEMKKETEDMCLTDHQPLDFSRVPGLTLPAGAFSDCL 828
Db 1845 -----EQKQTVATSTTSPTSSTST----- 1866
Qy 839 TIVEFLHSFGKVLGPDPAKDVPSGLVQEGLLCQGDLSGEVQDLDLVRLLKAALHDPGPPS 888
Db 1867 -----ISPAQKVMVAPI--SGSVTTGTGM----- 1901
Qy 889 YCQSLKILGEKVSSEPLFRDNVSEILRCFLMAYKXPALCDRLRTPQFAQPPQQAVAL 948
Db 1902 PQQKNFHTFATWYKQGSN--SGVVQVQKVLGHIPSTGTSQ--QTFTSQP--RTATVT 1958
Qy 949 APFVHELKGTLLIINEIDKLTLESMSYRKNKWIVEGRLLR-----LKTVLAKRT--GRS-- 1000
Db 1959 IRPNTSGSGGT-----TSNQVITGPQIRPGMTVIRPQLQOSTLGAII 2002
Qy 1001 ----EVENGRPECLGRRRSRIMBETSGMBEEEEESIAAVPGR-----GRRDGEVDAT 1052
Db 2003 RTPVMVQGAQVQV-----TQIRGQPVSTAVSAPNTVSTPGQKSLTSATSTSIQSS 2057
Qy 1053 ASSIPELERQIEKLSKRLFRKLLHSSQMLRAVSLQODRYRRRYWVLPYLAGIFVEGT 1112
Db 2058 ASQPPRQOGVQKLTMAQLTQLTGHGNGQGLTVVIQGGTGLQLIP--QGVTVLPG 2115
Qy 1113 EGNLVPEVIEKTDLSLVAHAASLNPALFSMKMELAGSNNTASSPARARSLPKTKPGF 1172
Db 2116 PQQJMQAMNGT-----VQRELFIT--PLATTATASTTTTTSV--TTAAG-- 2158
Qy 1173 MQPRHFKSPVRGQDSEQPAQLQPAQLHVPAPQOPQLQLOLQSHKFLQEGSPSLSGQ 1232
Db 2159 -----TGEQRQSKLSPQMVGHDKTLPP----- 2181
Qy 1233 SOHDLQSALFSLWSLQTSQSHSLSSSVLTPDPSGKLDPAQSPPEPEPEDEAESDPL 1292
Db 2182 -----AQSSSVGAFAEAQPTAQPSAQPOPTQPSQPAQVEQTOPEV 2223
Qy 1293 QAFWENISAQPCNAAPTPPLAVSBDQTPSPQQLASSKPMNRPSAANPCSPVQFSSNPL 1352
Db 2224 QT-QTTVSHVPSEAQPT-----HAQSKPQVAAQSQOSQVNOVGQSPVRVQSPQTRI 2275
Qy 1353 AGLAPKRAGDPGEMPPQSPFTGLGP-----KRRGRPPSKFFKQMBQRYLTQLTA 1401
Db 2276 RPSTPSQLS--PGQSQVQTTTSQPIPIQHTSLQIPSQGQPSQSPQVVMKXNAVIEHLK 2333
Qy 1402 QP---VPPEMCSGWVWP-----DPEMLDAMLKALHPRGIREK-----ALH 1439
Db 2334 QKSKMTPAERENQRMVNCNMVYILDKIDKEEQAAKGRBESVQKRSKQNAIKLS 2393
Qy 1440 KHLNKHDFLOEVCVLRPSADPIFERQLPAQFEGIMSMSPKEKTYETDLAVLQWVEELEQ 1499
Db 2394 ALLFKHKQLRAELIK-----KRALDKDLQI-EVQELKR 2428
Qy 1500 RVIMSDLOIRGWTCPSPDSTREDLAYCEHLSQSDITWRGPGREGLAPQRKTNPLDLA 1559
Db 2429 -----DLKIK-----KEKDLMLQAATAVAACPVPVAPP-----APPAPPSP----- 2468
Qy 1560 VMLAALQONKRRYLRBLPMPHTEVLEKALLSTPN--GAPEGTTTISVEITPRIR-- 1615
Db 2469 -----PPPPAVQHTGLLSTPTLPAASQKRKRREKXDSKSKKK 2507
Qy 1616 --INRQTLQRCRSAHVCLGLHLERS-----IAW-----EKSVMKVTCLVC 1655
Db 2508 KMISTTSKETKDTKLYCICTKTPYDESKFYIGCDLCTNWHGECVGITEREAKMDVVIC 2567
Qy 1656 -----RKGNDEFLLLCD-----GCDRGCHYICH-----RPRMEAVPEGDWF 1692
Db 2568 NDCKRAQSGSSEELYICRTPYDESQFYIGCDR--QNNWYHRCVGLQSGAEELIDE--YV 2624

Qy 1693 CTVCLAQOGEFTQKPGFKRGOKRSGYSLNFSEGDGRRERRVLLKGRESPAAGPRYSE 1752
Db 2625 CPQC--QSTEDAMTV----- 2637
Qy 1753 ERLSPSKRRRLSLSMENHSDLTFCBIIIMEMESHDAAMPFXEPVNPRLVSGYRRRIKPNMD 1812
Db 2638 --LTPLTEK-----DYEGLKRVLSLQAHKWAMPFLEPVPDNDAPDYGVVKEPMD 2686
Qy 1813 FSTWRELLRGYTSSEBFAADALLVFDNCOTFNEDDSEVGKAGHIMRRFPFESWBEFYQ 1872
Db 2687 LATMEERVQRYEYKLETFEVDMTKIFDNCRYNPSDSPFYQCAEVLESFFVQKLKGFKA 2746
Qy 1873 GKQAN 1877
Db 2747 SRSHN 2751
RESULT 35
AAB93746
ID AAB93746 standard; protein; 475 AA.
XX AAB93746;
XX AC
XX DT 26-JUN-2001 (first entry)
XX Human protein sequence SEQ ID NO:13405.
DE Human; primer; detection; diagnosis; antisense therapy; gene therapy.
KW Homo sapiens.
OS Homo sapiens.
FN EPI074617-A2.
XX 07-FEB-2001.
XX 28-JUL-2000; 2000EP-00116126.
XX 29-JUL-1999; 99JP-00248036.
PR 27-AUG-1999; 99JP-00300253.
PR 11-JAN-2000; 2000JP-00118776.
PR 02-MAY-2000; 2000JP-00183767.
PR 09-JUN-2000; 2000JP-00241899.
XX (HELI-) HELIX RES INST.
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI; 2001-318749/34.
DR
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.
XX
XX Claim 8; SEQ ID NO 13405; 2537pp + Sequence Listing; English.
XX
XX The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the

```
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893
CC represent human amino acid sequences; and AAH13629 to AAH13632 represent
CC oligonucleotides, all of which are used in the exemplification of the
CC present invention
XX
SQ Sequence 475 AA;

Query Match 3.9%; Score 379; DB 4; Length 475;
Best Local Similarity 23.1%; Pred. No. 1.7e-16;
Matches 109; Conservative 58; Mismatches 109; Indels 196; Gaps 9;

QY 1561 MRLAAEQNKRRYLREPL-----WPTHEVVLKALLSTPNGAPEGTTTETSYEITPRIR 1615
DQ 1 MALFQIEQGIERRFLKAPLDASDSGRSYKTVLDR----- 34

QY 1616 IWRQTLQRCSAAHVCLICGHLERSIAWEKSVNKVTCVLCVRKGNDEFLLLLCDGDRGCH 1675
DQ 35 -WRESLLSSASLSQVFLHLSTLDRSVIWSKSIILNARCKICRKGDAENNVLCDCGDRGHH 93

QY 1676 IYCHRPKMEVPEGDWFTCTVCLAQO-----VEGEFTQKPGPK 1713
DQ 94 TYCVRPKLKTVPEGWDFCEPCRPQRRLSRQPSLESDVEDSMGGDEDDVDGDEE 153

QY 1714 RGQKKSGYSINFSGD-----GR---RRRVLLKGR-----ESPAAG 1747
DQ 154 EQQSEEEYVEQDEDDQSEEEVSLPKRGPOVFLPKVTRGKLSFSSFGQQQEPGRY 213

QY 1748 PRYSE-----ERLSP 1757
DQ 214 PSRSQQSTPKTTSVSKTGRSLRKINSAPPTETKSLRIASRSTRSHGSHPLQADVVELLSP 273

QY 1758 SKRR-----LSMRNHS----- 1770
DQ 274 RKRRGRKSANNTPENSPNFRVIAAKSEQSRSVNIASLSQESSESKRCRCKRQSP 333

QY 1771 -----DLTFCEIILMESHDAAPFPXEPVNPRLVSGVRRITKPM 1811
DQ 334 EPSPTVTLGRSSGROGGVHLSAFQLVVELVRHDDSNPFLKLSKIQVPDYDIKKPI 393

QY 1812 DFTWMRELLRGYTSSEFAADALLVFDNCQTFNEDDSYGVKAGHIMRRFF 1863
DQ 394 ALNIIREKVNKCEYKLASEFIDDIELMFSCNCFEYNPRNTSEAKAGTLQAF 445

RESULT 36
AAU16212
XX AAU16212 standard; protein; 160 AA.
AC AAU16212;
XX
DT 07-NOV-2001 (first entry)
DE Human novel secreted protein, Seq ID 1165.
KW Human; immunosuppressive; antiarthritic; antirheumatic; cytostatic;
KW cardiac; vasotropic; cerebroprotective; nootropic; neuroprotective;
KW antibacterial; virucide; fungicide; ophthalmological; vulnerary;
KW secreted protein; rheumatoid arthritis; hyperproliferative disorder;
KW cardiovascular disorder; cardiac arrest; cerebrovascular disorder;
KW cerebral ischaemia; angiogenesis; nervous system disorder;
KW Alzheimer's disease; infection; ocular disorder; corneal infection;
KW wound healing; epithelial cell proliferation; skin ageing; food additive;
KW preservative; antiproliferative.
XX
OS Homo sapiens.
XX
FN WO20015322-A2.
XX
PD 02-AUG-2001.
XX
```

```
PF 17-JAN-2001; 2001WO-US001341.
XX 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234977P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
```

PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 06-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251988P.
PR 11-DEC-2000; 2000US-0251990P.
PR 05-JAN-2001; 2001US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
PR (HUMA-) HUMAN GENOME SCI INC.
XX
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-488783/53.
XX N-PSDB; AAS26199.
XX

PT New nucleic acid molecules encoding 461 human secreted proteins for
PT diagnosing, preventing, treating or ameliorating medical conditions and
PT used as food additives or preservatives.
XX
XX Claim 11; SEQ ID NO 1165; 980pp; English.
XX
CC The invention relates to isolated nucleic acid molecules and their
CC encoded secreted proteins. The nucleic acids and proteins are used to
CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,
CC rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used
CC in diagnosing a pathological condition or susceptibility to a
CC pathological condition. Antibodies to the proteins can also be used in
CC alleviating symptoms associated with the disorders and in diagnostic
CC immunoassays e.g. radioimmunoassays or enzyme linked immunosorbent assays
CC (ELISA). Disorders which are diagnosed or treated include autoimmune
CC diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g.
CC neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac
CC arrest, cerebrovascular disorders e.g. cerebral ischaemia, angiogenesis,
CC nervous system disorders e.g. Alzheimer's disease, infections caused by
CC bacteria, viruses and fungi and ocular disorders e.g. corneal infection,
CC and many other disorders listed in the specification. The polypeptides
CC can also be used to aid wound healing and epithelial cell proliferation,
CC to prevent skin aging due to sunburn, to maintain organs before
CC transplantation, for supporting cell culture of primary tissues, to
CC regenerate tissues and in chemotaxis. The polypeptides can also be used
CC as a food additive or preservative to increase or decrease storage
CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,
CC minerals, cofactors and other nutritional components. The present
CC sequence represents a novel secreted protein of the invention. Note: The
CC sequence data for this patent did not form part of the printed
Query Match 3.6%; Score 351; DB 4; Length 160;
Best Local Similarity 41.6%; Pred. No. 2.5e-15;
Matches 69; Conservative 29; Mismatches 36; Indels 32; Gaps 3;
QY 1713 KRG-----QKRKSGYSLNFSFGDGRRRVLLKGSPAGPRYSERLSPSKRRRLSMR 1765
DB 19 KRGKNDLKKRWENTSLNSK-----QSFTSVKPK---R 52
QY 1767 NHHSDLTFCEIILMEMESHDAWPFBPVNPRLVSGYRRIIKNPMDFSTMRLLRGY 1826
DB 53 DDSKDLCALCSMLTETMETHEDAWPFLPLVNLKLVGYKKVKKPMDFSTIREKLSSGQYP 112
QY 1827 SSEEFAADALIVDNCOTENEDDSEVGKAGHIMRFFESRWEFFYQ 1872
DB 113 NLETFALDVLVDFDNCETFNEDDSDIGRAGHNMKRYFEKKWTDTPK 158
RESULT 37
ABU55281
ID ABU55281 standard; protein; 160 AA.
XX
XX AC ABU55281;
XX
XX 18-MAR-2003 (first entry)
XX
XX Human novel polypeptide #368.
DE
XX
XX Human; neural disorder; immune system disorder; renal disorder;
KW muscular disorder; respiratory disease; reproductive disorder;
KW gastrointestinal disorder; pulmonary disorder; cardiovascular disorder;
KW hyperproliferative disorder; inflammatory disease; allergic reaction;
KW blood related disorder; cancer; immunosuppressive; antiinflammatory;
KW cardiovascular; nephrotropic; cytostatic; antiallergic; thrombolytic;
KW haemostatic; antiarteriosclerotic.
XX
OS Homo sapiens.
XX
XX US2002132753-A1.
XX
XX 19-SEP-2002.
XX
XX 17-JAN-2001; 2001US-00764864.
XX

XX 31-JAN-2000; 2000US-0179065P.
 PR 04-FEB-2000; 2000US-0180628P.
 PR 28-JUN-2000; 2000US-0214886P.
 PR 07-JUL-2000; 2000US-0216647P.
 PR 07-JUL-2000; 2000US-0216880P.
 PR 11-JUL-2000; 2000US-0217487P.
 PR 11-JUL-2000; 2000US-0217496P.
 PR 14-JUL-2000; 2000US-0218290P.
 PR 26-JUL-2000; 2000US-0220963P.
 PR 26-JUL-2000; 2000US-0220964P.
 PR 14-AUG-2000; 2000US-0224518P.
 PR 14-AUG-2000; 2000US-0224519P.
 PR 14-AUG-2000; 2000US-0225267P.
 PR 14-AUG-2000; 2000US-0225268P.
 PR 14-AUG-2000; 2000US-0225270P.
 PR 14-AUG-2000; 2000US-0225447P.
 PR 14-AUG-2000; 2000US-0225757P.
 PR 14-AUG-2000; 2000US-0225758P.
 PR 22-AUG-2000; 2000US-0226868P.
 PR 30-AUG-2000; 2000US-0228924P.
 PR 01-SEP-2000; 2000US-0229287P.
 PR 01-SEP-2000; 2000US-0229343P.
 PR 01-SEP-2000; 2000US-0229344P.
 PR 01-SEP-2000; 2000US-0229345P.
 PR 05-SEP-2000; 2000US-0229509P.
 PR 05-SEP-2000; 2000US-0229513P.
 PR 08-SEP-2000; 2000US-0231413P.
 PR 21-SEP-2000; 2000US-0234223P.
 PR 21-SEP-2000; 2000US-0234274P.
 PR 25-SEP-2000; 2000US-0234997P.
 PR 27-SEP-2000; 2000US-0235834P.
 PR 29-SEP-2000; 2000US-0236327P.
 PR 29-SEP-2000; 2000US-0236367P.
 PR 29-SEP-2000; 2000US-0236368P.
 PR 29-SEP-2000; 2000US-0236369P.
 PR 29-SEP-2000; 2000US-0236370P.
 PR 02-OCT-2000; 2000US-0236802P.
 PR 02-OCT-2000; 2000US-0237037P.
 PR 02-OCT-2000; 2000US-0237038P.
 PR 02-OCT-2000; 2000US-0237039P.
 PR 02-OCT-2000; 2000US-0237040P.
 PR 13-OCT-2000; 2000US-0239935P.
 PR 20-OCT-2000; 2000US-0240960P.
 PR 20-OCT-2000; 2000US-0241785P.
 PR 20-OCT-2000; 2000US-0241809P.
 PR 01-NOV-2000; 2000US-0244617P.
 PR 17-NOV-2000; 2000US-0249299P.
 PR 08-DEC-2000; 2000US-0251856P.
 PR 08-DEC-2000; 2000US-0251868P.
 PR 08-DEC-2000; 2000US-0251869P.
 XX (ROSE/) ROSEN C A.
 PA (RUBE/) RUBEN S M.
 PA (BARA/) BARASH S C.
 XX Rosen CA, Ruben SM, Barash SC;
 XX WPI; 2003-147444/14.
 DR N-PSDB; ABX73540.
 XX
 PT New polypeptides and nucleic acids, useful in gene therapy for treating,
 PT inhibiting or preventing e.g. neural, immune system, muscular,
 PT respiratory, reproductive, gastrointestinal, pulmonary, cardiovascular or
 PT renal disorders.
 XX
 PS Claim 11; SEQ ID NO 1165; 402pp; English.
 XX
 CC The invention relates to human novel polypeptides and their associated
 CC polynucleotides. The polypeptides and polynucleotides are useful in gene
 CC therapy for treating, inhibiting or preventing neural disorders, immune
 CC system disorders (e.g. systemic lupus erythematosus, rheumatoid arthritis
 CC and multiple sclerosis), muscular disorders, respiratory diseases (e.g.

CC nasal vestibulitis, nasal polyps and sinusitis), reproductive disorders,
 CC gastrointestinal disorders, pulmonary disorders, cardiovascular disorders
 CC (e.g. congenital heart defects, Ebstein's anomaly and hypoplastic left
 CC heart syndrome), renal disorders (e.g. acute kidney failure and end-stage
 CC renal disease), hyperproliferative disorders (e.g. Hodgkin's disease and
 CC leukaemia), inflammatory diseases (e.g. septic shock, bursitis and
 CC appendicitis), allergic reactions and conditions (e.g. asthma), blood
 CC related disorders (e.g. thrombosis, atherosclerosis and myocardial
 CC infarction) and cancerous diseases. Sequences ABU54914-ABU55699 and
 CC ABU55748 represent human novel polypeptides of the invention
 XX
 SQ Sequence 160 AA;
 Query Match 3.6%; Score 351; DB 6; Length 160;
 Best Local Similarity 41.6%; Pred. NO. 2.5e-15;
 Matches 69; Conservative 29; Mismatches 36; Indels 32; Gaps 3;
 QY 1713 KRG-----QKRGSGYSLNFSEGDGRRRRVLLKGRSPAAGPRYSEERLSPSKRRRLSNR 1766
 Db 19 KRGKNDLKKRQVEENTSINLSK-----QESFTSVKKPK---R 52
 QY 1767 NHHSDLTPECEILMEMESHDAAPPEXEPVNPRLVSGYRRIIKNPMDFGTMRELLRGVY 1826
 Db 53 DSKDLALCSMLTETHEDAWPFLLPVNLKLVPGYKVKVKKPMDFSTIREKLSGGYP 112
 QY 1827 SSEEFAADALVFDNCOTFNEDDSVKGAGHIMRRFFESRWEEFYQ 1872
 Db 113 NLETFALDVLVFDNCETFNEDDSDIGRAGHMRKYFEKKWTDTPK 158
 RESULT 38
 AAM93902
 ID AAM93902 standard; protein; 1050 AA.
 XX
 AC AAM93902;
 XX
 DT 06-NOV-2001 (first entry)
 XX
 DE Human polypeptide, SEQ ID NO: 4045.
 XX
 KW Human; full length cDNA; cDNA synthesis; oligo-capping.
 XX
 OS Homo sapiens.
 XX
 PN EP1130094-A2.
 XX
 PD 05-SEP-2001.
 XX
 PF 07-JUL-2000; 2000EP-00114089.
 XX
 PR 08-JUL-1999; 99JP-00194486.
 PR 11-JAN-2000; 2000JP-00118774.
 PR 02-MAY-2000; 2000JP-00183765.
 XX
 XX (HELI-) HELIX RES INST.
 XX
 PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
 PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
 XX
 DR WPI; 2001-524255/58.
 DR N-PSDB; AAK94864.
 XX
 PT 830 Primers useful for synthesizing full length cDNA clones and their use
 PT in genetic manipulation.
 XX
 PS Claim 8; SEQ ID NO 4045; 1380pp + Sequence Listing; English.
 XX
 CC The invention relates to primers for synthesising full length cDNA
 CC clones. 830 cDNA molecules encoding a human protein have been isolated
 CC and nucleotide sequences of 5'- and 3'-ends of the cDNA molecules have
 CC been determined. Primers for synthesising the full length cDNA are useful
 CC for clarifying the function of the protein encoded by the cDNA. The full
 CC length clones were obtained by construction of full length enriched cDNA

CC libraries that were synthesised by the oligo-capping method. The primers
CC enable the production of the full length cDNA easily without any special
CC methods. The present sequence is a polypeptide encoded by a full length
CC human cDNA of the invention. Note: the sequence data for this patent did
CC not form part of the printed specification, but was obtained in CD-ROM
CC format directly from EPO
XX
SQ Sequence 1050 AA;

```
Query Match          3.4%; Score 338; DB 4; Length 1050;
Best Local Similarity 20.7%; Pred. No. 3.3e-13;
Matches 204; Conservative 125; Mismatches 335; Indels 322; Gaps 45;

Qy 974 SYRKNKWTVEGRRLRL--KTVLAKRTGRSEVEMGRPEECIGRRSSRIMEETSGMEEREE 1031
Db 262 AFQNKVIIDTLITKMEKTKYIKFT-----GNQIQNRILIEVQNQKQVEQ 307

Qy 1032 BESTAAMP-----GRRGRDGEVDATASSIPELERQIEKLSKQQLFFRKLHSSQMLR- 1085
Db 308 DIKVAIFTLMVEINKKGK-----ALLHQLSLAKDH---RMKLMQQQOEAVG 351

Qy 1086 -----AVSLGQDR---YRRYVWLPY-LAGIFVETGEGNLVPEEVIKKETD 1127
Db 352 LSKOLEHVMHFSKWAVSSGSSALLYSKR--LITYRLRHLRLRARCDSPTVNTNTIQFHC 409

Qy 1128 SLKVAHAASLNPALFSKMWELAGSNTTASSPARARSRLTKPGFMOPRHPKSPVRGDS 1187
Db 410 -----PSFWAQINILGSLVIEDK-----ESQFQMPK-----QNPVVEQNS 445

Qy 1188 EQPQA-----QLQP-EAQLHVPAQPOQLQLQSHKGFLEQEGSPLSL--QQSQHDLQS 1240
Db 446 QPPSGLSSNQLSKFPTQLSLAQLRLQHQVMAQRQVQRRPAPVGLPNPQMGPQIQP 505

Qy 1241 AFLSWLSQTHSSLLSSSVLTPDSSPGKLDPAQSQ-----PPEEPEDAESSPDQAQFW 1297
Db 506 S-----ISHQQPPRLINQNHSPKNGPVLPPHPQQLRYPNPQNIPOAIKPNPLQMAFL 561

Qy 1298 NISAQMCNAAPTPLAVSEDOPTPSPOOLASSKMPNRPNSAANC-----SPVQFSS 1349
Db 562 AQQA-----IKWQIISGGQTPSTNTSSSTP--SSPTITSAGYDGKAFGFMIDLS 612

Qy 1350 TPLAGLAPKRAGDPGEMQSP-----TGL--GQPKRRGRPPSKFPKQ 1390
Db 613 SPVGG-----SYNPLSLPDIDCSITMLDNIVKDTNIDHCQ-----RPSNRITVQ 659

Qy 1391 MEQRYLTQTAQVPPPECSGWWIPDPPEMLDAM-LKALHPRGIREKALHKLHKNHRDFL 1449
Db 660 SPNSS-----VPSPLAGPVTMTSVHP----- 681

Qy 1450 QEVCLRPSADPIFEPRQLPAFQEGIMSWPK-----EKTVEYDLAVLOWV-----E 1495
Db 682 -----PIRSPASSVSGRSSGSSKAPAGADSTHKVPVWLEFIRIKQENSGPPE 731

Qy 1496 ELEQRVIM-----SDLIQIRGWTCPSPDSTREDLAYCEHLSDSQEDITWPGREGIAPQK 1551
Db 732 NYDPFVIVQESDEESRPNQANYPRSLTSLSSNSQSSSTSEENVLRSD-----APDST 786

Qy 1552 TTNPLDLAVMLALEQNVKRYRLREPLWPTHEVVLEKALLSTNGAPGTTTISYEIT 1611
Db 787 GDQP-----GLHQD-----NSSNGKSE----- 803

Qy 1612 PRIIRWQTLQRCSSAAHVCILGHLERSIAWEKSVNKTCLVCRKGDNDSEFLALCDGCD 1671
Db 804 -----WLDPSQ--KSPLHV-----GETRK-----EDPDNEDWCAVCQNGE---LICCCKP 845

Qy 1672 RGCHYICHRPKMEAVPEGDWFTCTVLAQQVGEFTQKPFKRGOKRSGYSLSNPFSEGDG 1731
Db 846 KVFHLSCHVPTLTNPPSEWICTFC-RDLSKPEVEYDCDASHNSEKKK-----TEG-- 896

Qy 1732 RRRVLLKGRSPAGPRYSBERLSPSKRRLSMRNHHSDITFCRILMESHESDAWPF 1791
Db 897 -----LVK-----LTPIDKRK-----CERLLLFYCHEMSLAF 924
```

```
Qy 1792 XEPVNPRLVGYRRIIKNPMDFSMTBRLLR--GGYTSSSEFAADALIVFNCQTFNEDD 1849
Db 925 QDPV-PLTVDPYKILIKNPMDLSIIKKRLQEDYSWYKPEDFVADFRILFQNCFAEFNEPD 983

Qy 1850 SEVGKAGHIMRRFPFESRWEFYOQKQ 1875
Db 984 SEVANAGIKLENYFEELLKNLYPEKR 1009

RESULT 39
ADB80954
ID ADB80954 standard; protein; 1050 AA.
XX
AC ADB80954;
XX
DT 04-DEC-2003 (first entry)
XX
DE RING-SH complex related protein, SEQ ID No 28.
XX
KW RING-SH 3; Gag protein; Gag late domain; P13K; actin; myosin; Hsp60;
KW Hsp70; Hsp90; STAM1; STAM2A; STAM2B; VHS-UIM; GTPase; E2 enzyme; tag101;
KW cullin; RING-SH; clathrin; virucide; vaccine; antiviral; retrovirus;
KW rhabdovirus; filovirus.
XX
OS Unidentified.
XX
PN WO2003033646-A2.
XX
PD 24-APR-2003.
XX
PF 31-JUL-2002; 2002WO-US024589.
XX
PR 31-JUL-2001; 2001US-0308958P.
XX
PR 09-NOV-2001; 2001US-0345846P.
XX
PA (PROT-) PROTEOLOGICS INC.
XX
PI Greener T, Moskowitz H, Reiss Y, Alroy I;
XX
XX WPI; 2003-393509/37.
DR N-ESDB; ADB80995.
XX
XX New isolated protein complex comprising a RING-SH 3 polypeptide and
PT another polypeptide, useful for detecting cells infected with a virus,
PT and for treating viral disorders caused by retroviruses, rhabdoviruses,
PT or filoviruses.
XX
PS Disclosure; Fig 28; 176pp; English.
XX
XX The invention relates to a novel isolated protein complex comprising a
CC RING-SH 3 polypeptide and a polypeptide selected from a Gag protein, a
CC Gag late domain, P13K, actin, myosin, Hsp60, Hsp70, Hsp90, STAM1, STAM2A,
CC STAM2B, VHS-UIM, a GTPase, an E2 enzyme, tag101, a cullin, RING-SH, and a
CC clathrin. The novel protein complex has virucide activity and can be used
CC to treat disorders as part of a vaccine. The protein complex and
CC composition are useful for detecting cells infected with a virus, for
CC identifying agents having antiviral activity, and for treating viral
CC disorders caused by retroviruses, rhabdoviruses, or filoviruses. This
CC sequence is a protein comprising the RING-SH complex of the invention.
XX
SQ Sequence 1050 AA;
```

```
Query Match          3.4%; Score 338; DB 7; Length 1050;
Best Local Similarity 20.7%; Pred. No. 3.3e-13;
Matches 204; Conservative 125; Mismatches 335; Indels 322; Gaps 45;

Qy 974 SYRKNKWTVEGRRLRL--KTVLAKRTGRSEVEMGRPEECIGRRSSRIMEETSGMEEREE 1031
Db 262 AFQNKVIIDTLITKMEKTKYIKFT-----GNQIQNRILIEVQNQKQVEQ 307

Qy 1032 BESTAAMP-----GRRGRDGEVDATASSIPELERQIEKLSKQQLFFRKLHSSQMLR- 1085
Db 308 DIKVAIFTLMVEINKKGK-----ALLHQLSLAKDH---RMKLMQQQOEAVG 351
```


Qy 1086 -----AVSLGQDR---YRRYVWLPY-LAGIFVETGEGNLVPEEVIKKTD 1127
Db 352 LSKOLEHVMHFSKWAVSGSSTALLYSKR--LITVRLHLLRARCDASPTVNTTIQFHCD 409
Qy 1128 SLKVAHAASLNPALPSMKWELAGSNNTASSPARASRLPKTKPGFMQPRHFKSPVRGQDS 1187
Db 410 -----PSFWAQNIINLGSVIEDK-----ESQPMKP-----QNPVVEQNS 445
Qy 1188 EQPOA----QLQP-EAQLHVPAPQPOQLQLOSHKGFLEQEGSPLSL--GOSQHDLSQS 1240
Db 446 QPPSGLSNQLSKFPTQISLAQLRLQHMQQVMAQROQVRRPAPVGLPNRPMQGPQIQP 505
Qy 1241 AFLSWLSQTSQSHSSLLSSSVLTPSSPGKLDPAFSQ---PPEEPEDAEASSPDQAQFWF 1297
Db 506 S-----ISHQQPPRLINFNHSPKPNGLVPPHPPQQLRYPPNQNIPIROAIKPNPLQMAFL 561
Qy 1298 NISQMPCNAAPTPLAVSEDOPTPSPOQLASSKPMNRPSAANPC-----SPVQFSS 1349
Db 562 AQQA-----IKQWQISSCGGTPTNSTSTP--SSFTTISAAGYDGAFGSPMIDLS 612
Qy 1350 TPLAGLAPKRRAGDPGEMPQSP-----TGL--GQPKRRGRPPSKFKQ 1390
Db 613 SPVGS-----SYNLPSPDIDCSSTIMLDNIVRKDTNIDHQP----RPSNRIVQ 659
Qy 1391 MEQRYLTQLTAPVPPCMCGMWIIPDPEMLDAM-LKALHPRGIREKALHKLNRHDFL 1449
Db 660 SPNSS-----VPSPLAGPVTMTSVHP----- 681
Qy 1450 QEVCLRPSADIFEPRLQPAQEGIMSWSPK-----EKTYETDLAVLQWV-----E 1495
Db 682 -----PIRSPASSVSGSGSSGSSKPGAGADSTHKVPVVMLEPIRIKQENSGPPE 731
Qy 1496 ELEQRVIM----SDLOIRGWTCPSDSTREDLAYCEHLSDSQEDITWRGPGREGAPQK 1551
Db 732 NYDFPVVIVKQESDESPQANYPRSLTSLLLNSSQSSTSEEIVLRSD-----APDST 786
Qy 1552 TTNPLDLAVMLAALQNVKRYRLREPLWPTHEVVLEKALLSTPNAGPEGTTTISYEIT 1611
Db 787 GDQP-----GLHQD-----NSSNGKSE----- 803
Qy 1612 PRIRIWRQTLQRCSAAHVCLGLHLERSIAWEKSVNKTCLVCRKGDNDPFLLCDCGD 1671
Db 804 -----WLDPSQ--KSPLHV-----GETRK-----EDDPNEDWCAVCQNGE---LLCCEKCP 845
Qy 1672 RGCHLYCHRPKWEAVEPGEWDFCTVCLAAQVGEFTQKPGFKRGOKRSGYSLNFSQDG 1731
Db 846 KVFHLSCHVPLTNPSPGEWICTFC-RDLSKPEVEYDCDAFSHNSEKKK-----TEG-- 896
Qy 1732 RRRVLLKGRESAPGRYSEERLSPSKRRRLSMRHHSDLTFCBIIILMEMESHDAWPF 1791
Db 897 -----LVK-----LTPIDKR-----CERLLFLYCHEMSLAF 924
Qy 1792 XEPVNPRLVGYRRIIKNPMDFSTWRRLR--GGYTSSEFAADALLVFNQOTFNEDD 1849
Db 925 QDPV-PLTVDPYKIKIKNPMIDSLITKRLQEDYSWYKPDVADFRILFQNCAEFNEDP 983
Qy 1850 SEVGKAGHMRFPFESRWEFFVQKQ 1875
Db 984 SEVANAGIKLENYFEBLLKNLYPEKR 1009

RESULT 40
ADL32012
ID ADL32012 standard; protein; 1050 AA.
XX AC
XX ADL32012;
DT 20-MAY-2004 (first entry)
XX DE Human protein encoded by a full length cDNA clone SeqID 4045.
XX human; medicine; signal transduction; glycoprotein; transcription;
KW

oligo-capping method.
Homo sapiens.
EP1396543-A2.
10-MAR-2004.
07-JUL-2000; 2003EP-00025638.
08-JUL-1999; 99JP-00194486.
11-JAN-2000; 2000JP-00118774.
02-MAY-2000; 2000JP-00183865.
07-JUL-2000; 2000EP-00114089.
(REAS-) RES ASSOC BIOTECHNOLOGY.
Ota T, Nishikawa T, Isogai T, Hayaashi K, Ishii S, Kawai Y;
Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
WPI; 2004-204755/20.
N-PSDB; ADL32011.
New oligonucleotide primers (830 cDNAs) useful for synthesizing full
length human cDNAs.
Example 1; SEQ ID NO 4045; 1340pp; English.
This invention relates to a novel primers useful for synthesizing full
length cDNA molecules that encode human proteins. Specifically, it refers
to secretory or membrane proteins that are potential therapeutic agents/
target molecules in the field of medicine, and in particular genes
encoding proteins that are associated with signal transduction,
glycoproteins and transcription. The present invention describes a method
for efficiently cloning a full length human cDNA from both the 5' and 3'
ends using the oligo-capping method. This polypeptide sequence is a full
length human protein of the invention.
Sequence 1050 AA;
Query Match 3.4%; Score 338; DB 8; Length 1050;
Best Local Similarity 20.7%; Pred. No. 3.3e-13;
Matches 204; Conservative 125; Mismatches 335; Indels 322; Gaps 45;
Qy 974 SYRKNKVIQGLRRL--KTVLAKRTGRSEVMGRPECLGRRSSRIMESTSGMEBEE 1031
Db 262 AFQKQVLIIDTLTKMEKTKYIKFT-----CQIQNRRIIEVNQKQVEQ 307
Qy 1032 EESIAAVP-----GRRRRDGEVDATASSIPELERQIEKLSKQLFFRKLLHSSQMR- 1085
Db 308 DIKVAIFTLMVEINKKGK-----ALLHQLLESKADH---RMKLMQQQVEVAG 351
Qy 1086 -----AVSLGQDR---YRRYVWLPY-LAGIFVETGEGNLVPEEVIKKTD 1127
Db 352 LSKOLEHVMHFSKWAVSGSSTALLYSKR--LITVRLHLLRARCDASPTVNTTIQFHCD 409
Qy 1128 SLKVAHAASLNPALPSMKWELAGSNNTASSPARASRLPKTKPGFMQPRHFKSPVRGQDS 1187
Db 410 -----PSFWAQNIINLGSVIEDK-----ESQPMKP-----QNPVVEQNS 445
Qy 1188 EQPOA----QLQP-EAQLHVPAPQPOQLQLOSHKGFLEQEGSPLSL--GOSQHDLSQS 1240
Db 446 QPPSGLSNQLSKFPTQISLAQLRLQHMQQVMAQROQVRRPAPVGLPNRPMQGPQIQP 505
Qy 1241 AFLSWLSQTSQSHSSLLSSSVLTPSSPGKLDPAFSQ---PPEEPEDAEASSPDQAQFWF 1297
Db 506 S-----ISHQQPPRLINFNHSPKPNGLVPPHPPQQLRYPPNQNIPIROAIKPNPLQMAFL 561
Qy 1298 NISQMPCNAAPTPLAVSEDOPTPSPOQLASSKPMNRPSAANPC-----SPVQFSS 1349
Db 562 AQQA-----IKQWQISSCGGTPTNSTSTP--SSFTTISAAGYDGAFGSPMIDLS 612
Qy 1350 TPLAGLAPKRRAGDPGEMPQSP-----TGL--GQPKRRGRPPSKFKQ 1390

```
Db 613 SPVGG-----SYNLPSPFDIDCSSTIMLDNIVRDTNIDHQP-----RPSNRRTVQ 659
Qy 1391 MEQRYLTQTAQVPPEPCSGMWIPDPPEMLDAM-LKALHPRGIREKALHKLHKNHRDFL 1449
Db 660 SPNBS-----VPSGLAGPVTMTSVHP----- 681
Qy 1450 QEVCLRPSADPIFEFRQLPAFQEGIMSPK-----EKTETDLAVLOWV-----E 1495
Db 682 -----PIRPSASSVSRGSSGSSKPAGADSTHKVPVWLEPIRIKQENSGPPE 731
Qy 1496 ELEQRVIM-----SDIQIRGWTCPSPDSTREBLAYCEHLSDSQEDITWRGREGGLAPQRK 1551
Db 732 NYDFPVIVKQESDEESRPQANYPRSILTSILLNSQSSTSEETVLRS-----APDST 786
Qy 1552 TTNPDLDAVRLAALAEQNVKRYLREPLWPTHEVLEKALLSTNGAPEGTTTISYEIT 1611
Db 787 GDQP-----GLHOD-----NSNGKSE----- 803
Qy 1612 PRIIRWQTLQRCRAAHVCLGLHLERSIAWEKSVNKTCLVCRKGDNDFFLLCDGCD 1671
Db 804 -----WLDPSQ--KSPHAV-----GETRK-----EDDPNEDWCAVCQNGE---LLCCEKCP 845
Qy 1672 RGCHYCHRPKXEAPEGDWFCVTCLAAQOVGEFTQKPGPKRQGRKSGYSLNFSBGD 1731
Db 846 KVFHLSCHVPTLTNFPSSGEMICTFC-RDLSKPEVEYDCDAPSHNSEKK-----TEG-- 896
Qy 1732 RRRRVLLKGRSPAAGPRYSERLSPSKRRLSMRHHSDLTFCETIILMEMESHDAWPF 1791
Db 897 -----LVK-----LTPIDRK-----CERLLFLFYCHEMSLAF 924
Qy 1792 XEPVNPRLVSGYRRIIKNPMDFSTWRERLLR--GGYTSSEBFAADALIVFDCOTFNEDD 1849
Db 925 QDPV-PLTVDPYKLIKKNPMDLSTIKKQLQEDYSMYKPEDFVADFRLLIFQNCABFNEPD 983
Qy 1850 SEVGKAGHIMRRFESRWEEFYQKGQ 1875
Db 984 SEVANAGIKLENYFEELLKNLYPEKR 1009
```

RESULT 41
ADB80955
ID ADB80955 standard; protein; 1127 AA.
XX
AC ADB80955;
XX
DT 04-DEC-2003 (first entry)
XX
DE RING-SH complex related protein, SEQ ID No 29.
XX
KW RING-SH 3; Gag protein; Gag late domain; P13K; actin; myosin; Hap60;
KW Hsp70; Hsp90; STAM1; STAM2A; STAM2B; VHS-UIM; GTPase; E2 enzyme; tag101;
KW cullin; RING-SH; clathrin; virucide; vaccine; antiviral; retrovirus;
KW rhabdovirus; filovirus.
XX
OS Unidentified.
XX
PN WO2000303646-A2.
XX
PD 24-APR-2003.
XX
PF 31-JUL-2002; 2002WO-US024589.
XX
PR 31-JUL-2001; 2001US-0308958P.
PR 09-NOV-2001; 2001US-0345846P.
XX
FA (PROT-) PROTEOLOGICS INC.
XX
PI Greener T, Moskowitz H, Reiss Y, Alroy I;
PI WPI, 2003-393509/37.
DR N-PSDB; ADB80956.
XX

New isolated protein complex comprising a RING-SH 3 polypeptide and another polypeptide, useful for detecting cells infected with a virus, and for treating viral disorders caused by retroviruses, rhabdoviruses, or filoviruses.

Disclosure; Fig 29; 176pp; English.

The invention relates to a novel isolated protein complex comprising a RING-SH 3 polypeptide and a polypeptide selected from a Gag protein, a Gag late domain, P13K, actin, myosin, Hsp60, Hsp70, Hsp90, STAM1, STAM2A, STAM2B, VHS-UIM, a GTPase, an E2 enzyme, tag101, a cullin, RING-SH, and a clathrin. The novel protein complex has virucide activity and can be used to treat disorders as part of a vaccine. The protein complex and composition are useful for detecting cells infected with a virus, for identifying agents having antiviral activity, and for treating viral disorders caused by retroviruses, rhabdoviruses, or filoviruses. This sequence is a protein comprising the RING-SH complex of the invention.

Sequence 1127 AA;

Query Match 3.4%; Score 335; DB 7; Length 1127;
Best Local Similarity 19.5%; Pred. No. 5.8e-13;
Matches 241; Conservative 148; Mismatches 449; Indels 400; Gaps 52;

Qy 766 PACKADKTLATQRRLERQKQMIEMKKPTEDMCLTDHQPLDFGRVPLGLTLPAGFS 825
Db 118 PASLDDTCAVCQSQSLSRRAE---PKLLPCLHSFCL---RCLPEPERQLSVPIPGSGNG 171
Qy 826 D-----CLTIVEFLHSFGKVLGFPDPAKDVPSLGLVQEGLLCQ--DSLGE 868
Db 172 DIQGVGVIRCPVCRQECRQIDLVDNYFK-----DTSEAPSSSDEKSEQVCTCEDNAA 226
Qy 869 VQDLLVALLKAAALHDPGFPSPYCSQ--LKILGEKVSEIPLTRD-----NVSEILRCFLM 919
Db 227 V-----GFCVEGGEWLCKTCEAHQRVKFTKHLIRKEDVSESVG----- 267
Qy 920 AYGVPALCDRLRTQPFQAQFPQOKAAVLAPFVHELNGSTLIINEIDK-----TLESM 972
Db 268 ASGQRPVFC-----PVHKQEQKLFCETCDRLTCRDCQLLEHK 305
Qy 973 -----SSYRKKNWIVEGRRLRLKTVLAKRT-----GRSEVEMGRPECLGRRRSRIM 1020
Db 306 EHRYQFLEAFQKQGAENLLAKL---LEKNTVHFAATQVQ-----NRKIK 349
Qy 1021 B--ETSGMEEREEESIAAVFGRGRRDGEVDATASIPLEEROIEKLSKQLPFRKXLL 1078
Db 350 EVNETKRVQEIKAIFTL-----INEINKGKS---LQQLQEVTKER---QMKLL 396
Qy 1079 HSSQMLRAVSLGQDRYRRRYWVLPYLAGIFVEGTEGNLVPEE-----VIKKTDSL 1129
Db 397 QQQNDITGLS-----RQVKHVMNFTNWTWAIASGSSTALLYSKRLITFQLRHILKARCDPV 450
Qy 1130 KVAHAH---SLNPALFSWKMLAGSNNTASSPARASRPLKTKPGF----- 1172
Db 451 PAANGAIRFHCDPTFWAKNVVNLGNLVIESKPA-----FGYTPNVVVGVPVPGTN 500
Qy 1173 -----MQPRHFKSPVRGQDSEQQAQLQPEAQLHVPAPQOPQLQLQSHKG 1219
Db 501 HISKTPGQINLAURLQHMQQVYQKHQJQ-----QMRMQQPPAPVPTTTTQQH-- 553
Qy 1220 FLEQEGSPLSLQ-----SOHDLSSAFSLSMISQTSQSHSLSSSLTPDSSPG 1268
Db 554 --PRCAAPQMLQQQPPRLISVQWTQGRGNMNCGAF-----QAQMRLAQNAA---RIPG 601
Qy 1269 KLDPAQSPQPEEPDEAESPDLOAFWFNISAOQPCNAATPPLAVSEDOFTPSQOLA 1328
Db 602 -----IPRHSGFQYSMMQPHLQRHNS-----PGHAGFPVVSVHTTINPTPTTA 648
Qy 1329 SSKPMNRPSAANPCSPVQFSSTPLAGLAPKERRAGDPCGEMPOSPGLGQPKRRGRPP----- 1384
Db 649 TMANANR-----GPTSP-----SVTAIELIPS--VTNPENILPSLP-----DIPPIQLE 689
Qy 1385 ---SKFFKQMBQRYLTQTAQVPPEMCSGWWWIPTDPEMLDAMLKALHPRGIREKALHKH 1441

```
Db 690 DAGSSLDNLLSRY---ISGSHLPQPTSTMTMPSGSAISGSSGLS----- 734
Qy 1442 LNKHRDFLOEVLRLPSADIFPPQLPAPQEGIMWSPK--EKT---YETDLAVLQWVEE 1496
Db 735 -NSH-----TPVRPSTSTSGSRGSCSGSRTAKTSLSFKSDQVKYQBPFG 780
Qy 1497 LQORVIMSDLOIRGWTCPSPDSTREDLAYCEHLSQSDITWRGPGREGLAPQRTKTNPL 1556
Db 781 TEDEI-----CSFSGGVKQ-----EKTEGRRSACMLSSPESSLTPPLSTNLHL 824
Qy 1557 DLAVMLAALQONKRRYRLPLWPTHEVULEKALLSTPNGAPEGTTTETISVEITPRIRI 1616
Db 825 ESELDALASLENHVK-----TEPA-DMNESCKOSGLSLVNGK----- 861
Qy 1617 WRQTLQRCRSAHVCLGLHLERSIAWKSVMKVTCLVCRKGDNDFFLLCDGDRGCHI 1676
Db 862 -----SPIRSLMHSARIGDGN--KODDPNEDWCAVCONGD---LLCCEKCPKVFHL 911
Qy 1677 YCHPRMEAVPEGDFWCTVCLAQOQVEGFTQKPGPKRGQRKSGYSLNFSGDRRRRV 1736
Db 912 TCHVPTLLSFPFGDWICTFC----- 931
Qy 1737 LKKGRESPAAGPRYSEERLSPSKRRLSMRNHSDLTCEIILMEMESHDAWPFXEPVN 1796
Db 932 ----RDIGKPEVEYDCDNLQHSKKGTAQGLSPVDQKRCERLLLYLCHLSIEFQEPV- 986
Qy 1797 PRLVSGYRRIKNPDMDFSTMRLRLR--GGYTSSEEFADALLVFDNCOTFNE----- 1847
Db 987 PASIPNYKIILKPMDLSTVKKLQKSHQYQIPDDFVADVRLIFKNCERENMKVVQ 1046
Qy 1848 -----DDSEVGKAGHIMRFFSRWEEFYOGK 1874
Db 1047 VYADTQEINLKADSEVAQAGKAVLYFEDKLTETYSDR 1084

RESULT 42
ID AAB94613 standard; protein; 759 AA.
XX
AC AAB94613;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human protein sequence SEQ ID NO:15458.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-00116126.
XX
PR 29-JUL-1999; 99JP-00248036.
PR 27-AUG-1999; 99JP-00300253.
PR 11-JAN-2000; 2000JP-00118776.
PR 02-MAY-2000; 2000JP-00183767.
PR 09-JUN-2000; 2000JP-00241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-
PT length cDNAs defined in the specification, and for the detection and/or
PT diagnosis of the abnormality of the proteins encoded by the full-length
PT cDNAs.
```

```
XX Claim 8; SEQ ID NO 15458; 2537pp + Sequence Listing; English.
PS
XX
CC The present invention describes primer sets for synthesising 5602 full-
CC length cDNAs defined in the specification. Where a primer set comprises:
CC (a) an oligo-dT primer and an oligonucleotide complementary to the
CC complementary strand of a polynucleotide which comprises one of the 5602
CC nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in the
CC specification. The primer sets can be used in antisense therapy and in
CC gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893
CC represent human amino acid sequences; and AAH13629 to AAH13632 represent
CC oligonucleotides, all of which are used in the exemplification of the
CC present invention
XX
SQ Sequence 759 AA;
Query Match 3.2%; Score 316; DB 4; Length 759;
Best Local Similarity 20.1%; Pred. No. 6.1e-12;
Matches 158; Conservative 93; Mismatches 299; Indels 236; Gaps 28;
Qy 1150 GSNTTASSPARSRPLKTKPFQPRHFKSPVRGQDSEQPAQLQPEAQLHVPAPQPO 1209
Db 106 GTNHISKTPGOINLAQLRL-----QHMQQVVAQKHQQLQ-----QMRMQQPPAPVPT 153
Qy 1210 LQQLQSHKGFLEGGSPSLSGQ-----SQHDLQSASFLSWLSQTSQSHSLSS 1258
Db 154 TTTTQQH-----PRQAPQMLQQQPPRLISVQTMQRGNMNCAGF-----QAQMRLAQ 202
Qy 1259 SVLTPDSSPGKLDAPSQPPEPEPEDEAESPDLOAFWFNISAQMPCNAAPTPLAVSED 1318
Db 203 NAA---RIPG-----IPRHSGPQYSMMQPHLQROHSN-----PGHAGFPFVVSVHNT 246
Qy 1319 QPTSPQOLASSKPMNR-----PSAANP-----CSPVQFSSTPLA---- 1353
Db 247 TINTSPATMANARGPTSPSVTATELISVTNPENLPSLPDIPPIQANVPMHWSY 306
Qy 1354 GLAPKRAGDPGEMPQSPGTGLGQPKRGRPPSKFKQMEQRYLTQLTQAPVPPPCSGMW 1413
Db 307 EFGAREKTQDQNVLEDA-----GSSSLDNLSSRY---ISGSHLPQPTSTWN 350
Qy 1414 WIPDPEMLDAMLKALHPRGIREKALHKLNHGRDLOEVCLRPSADPIFEPRQLPAFOEG 1473
Db 351 PSPGFSALSPGSSGLS-----NSH-----TPVRPSTSTSGRG 384
Qy 1474 IMWSPK--EXT---YETDLAVLQWVELEQVIMSDLIQRGWTCPSDSTREDLAYCEH 1528
Db 385 SCGSGRTAETKTSLSFKSDQVKVQEPCTEDEI-----CSPSGGVKQ-----EK 428
Qy 1529 LSDSQEDITWRGPGREGLAPQRTKTNPLDLAVMLAALQONKRRYRLPLWPTHEVYLE 1588
Db 429 TEDGRRSACMLSSPESSLTPPLSTNLHLESLDALASLENHVK-----TEPA-DMNESCKQ 483
Qy 1589 KALLSTPNGAPEGTTTETISVEITPRIRIWRQTLQRCRSAHVCLGLHLERSIAWKSVM 1648
Db 484 SGLSLVNGK-----SPIRSLMHSARIGDGN--KDDDPN 518
Qy 1649 KVTCLVCRKGDNDFFLLCDGDRGCHIYCHRPKMEAVPEGDFWCTVCLAQOQVEGFTQK 1708
Db 519 EDWCAVCONGD---LLCCEKCPKVFHLTCHVPTLLSFPFGDWICTFC----- 563
Qy 1709 PGFPKRGQRKSGYSLNFSGDRRRRVLLKGRESPAAGPRYSEERLSPSKRRLSMRNH 1768
```


Qy	1044	RDGEVDATASSIPELEROIEKLSKROLFRFKKLLHSSQMLRAVSLGQDRY-----RRR	1097
Db	1670	-----AWEDTSPQONR-----YWRGR-----EDVALBQDTYWRLESCERK	1705
Qy	1098	YWVLPYLAG-----IFVEGTGNLVPEVIKKTDSLKVAHAASLNALFSMKWELAGSN	1152
Db	1706	VWFPHELDGQGARPHYTEERSTFLDEG---PDDQEVPLREHATRSWASDFK-----D	1757
Qy	1153	TTASSPARAR---SRPLKTKPGFMQRH-----FKSPVR-----GQD	1186
Db	1758	FOESSPOKGLEVERWLAESPVLGPPEEDKLTRSPFEIISPPASPPEMVQRPVSAPQOE	1817
Qy	1187	SEQQAOLQPEAQUH-----VPAQPQLOLQLOSH--	1217
Db	1818	SPIPDPKLPHMKNEPTTPSWLADIPWPVPKDRPLPAPLSPAPGPPTP-----APESHTP	1873
Qy	1218	-----KGFLEOGESPLS-LGOSQHDLS---QSAFLSWLSQTOSSHSL	1255
Db	1874	APFSWGTAEYDSVVAAVOEGAAELLEGFPYPLGKDYKKAERGEERABAPDKSHSK	1933
Qy	1256	LSSSVLTPDSGPKLDPAPSOPEPEPDPDAESSPDLOQAFWF-NISAOQ-----	1303
Db	1934	V-----PEASKSHATTEPEQ---TEPQREPTPYPDRSFQYADIYEQMLTLGLGPACPT	1985
Qy	1304	-----PC-----NAAPTPLAVLVSQDPT-----P	1322
Db	1986	REPLGAAGDWPPCLSTKEAAGAGNTSAEXELSSPISPKLSQDSTPTFSYAALGATVVP	2045
Qy	1323	SPOOLASKPMNRPSAANPCSPV---QFSSTPLAG---LAPKRAGDGPBMPQS-----	1370
Db	2046	RPEPGSMEPSLTPPAVPPRAPILSKGSPPLNGNILSCSPDRRSPKPKESGRSHWDST	2105
Qy	1371	-----FTGLOGPKRGRPSKPKOMEQRYLTQLTAAQVPPCMCSGWWIIPDEMLDAM	1424
Db	2106	SDELEKGAREQPEKEAQSPS-----PPHPIPMGSPTLW-PETE-----	2143
Qy	1425	LKALHPGRIEKALHKLHNKHRDPLQEVCLRPSADPIFEPRQLPAFOGIMSNSPKEKY	1484
Db	2144	---AHV---SPPLDSLHGP-----ARPSLD-----FPASAFGSSLOPAPP--	2178
Qy	1485	ETDLAVLQWVELEQRVIMSDLOIRGWTCPSPDSTREDLAYCEHLSQSDQBITWRQPGRE	1544
Db	2179	-----QLSPAPRPS---APCGSLAFSGDRALALAPG--	2207
Qy	1545	GLAPQKTTNPLDLAVMRLAALQONVKRYLRPLWPTHEVLEKALLSTP---NGAPEG	1601
Db	2208	---PPTTRHDEYLEVTKAPSLDSSLQ---LPSPSSPGAPLLSNLRPASPALSEGSSES	2262
Qy	1602	TTTTIYSIEITPRINWQTLQR-----CRSNAHVCLCL-----CHLERSATAWEK	1645
Db	2263	ATTPVISVAERFSPSLAEAEQSGELDPGWEPAAHSLWDLTFLSPAPPASLOLALAPAP	2322
Qy	1646	SVNKVTCLVCRKGNDFLLLLCDGCRGCHICY-----HRPKMEAVPEGDFWCTVCLAAQ	1700
Db	2323	SL-----PGDMGDIIL-----PCHLECSAATERPSPFQVPSED-----CAANG	2361
Qy	1701	VEGEFTQKPG-FPKRGOKRKSQGYSLNFSEG---DGRRRR-----VLLKGRE---SPAAG---	1747
Db	2362	PTETSPNPPGPAPAKAENEEAAACPAWEGAWPEGAERSRPDTLLSPSPQVCPAGGSGG	2421
Qy	1748	-----PRYSERLSPS	1758
Db	2422	PPSSASPEVEAGPQGCATEPRPHRGELSPS	2451

RESULT 44

AB858657

ID ABB58657 standard; protein; 2274 AA.

XX

AC ABB58657;

XX
DT 26-MAR-2002 (first entry)

XX	Drosophila melanogaster polypeptide SEQ ID NO 2763.
DE	
XX	Drosophila; developmental biology; cell signalling; insecticide;
KW	pharmaceutical.
KW	
XX	Drosophila melanogaster.
OS	
XX	WO200171042-A2.
PN	
XX	27-SEP-2001.
PD	
XX	23-MAR-2001; 2001WO-US009231.
XX	
XX	23-MAR-2000; 2000US-0191637P.
PR	
PR	11-JUL-2000; 2000US-00614150.
XX	
XX	(PEKE) PE CORP NY.
FA	
XX	
XX	Venter JC, Adams M, Li PWD, Myers EW;
PI	
XX	WPI; 2001-656860/75.
DR	N-PSDB; ABL02760.
XX	
DR	
XX	
PT	New isolated nucleic acid detection reagent for detecting 1000 or more
PT	genes from Drosophila and for elucidating cell signaling and cell-cell
PT	interactions.
XX	
XX	Disclosure; SEQ ID NO 2763; 21pp + Sequence Listing; English.
PS	
XX	
CC	The invention relates to an isolated nucleic acid detection reagent
CC	capable of detecting 1000 or more genes from Drosophila. The invention is
CC	useful in developmental biology and in elucidating cell signalling and
CC	cell-cell interactions in higher eukaryotes for the development of
CC	insecticides, therapeutics and pharmaceutical drugs. The invention
CC	discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC	sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC	ABB72072). The sequence data for this patent did not form part of the
CC	printed specification, but was obtained in electronic format directly
CC	from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX	
SQ	Sequence 2274 AA;
SQ	

Query Match.	3.1%;	Score 304;	DB 4;	Length 2274;
Best Local Similarity	20.0%;	Pred. No. 2.1e-10;		
Matches 325;	Conservative 184;	Mismatches 571;	Indels 544;	Gaps 68;
QY	15	TGLPPAPASGLKXPSSSCGELYT-----NGSPWNFPQOQKSLNGDVWVNGLSTVSHHTT	69	
Db	814	TTLPSTAASAL-----GLGTIIKGRSSPLNPLSAATSVGGAGI--GSSAVTGRSG	864	
QY	70	SGI---LNSAPHSSSTSHLHPPSVAYDCLWNYSQYPSANPSNLKOPPLLSQFSGGQYPL	126	
Db	865	AGVPVLIESGYSIE-----QSPKLMDRFRHHGLDS	896	
QY	127	NGILGSR-----QPS-----SPSHNTNLRAG-----SQKFWAN	155	
Db	897	DYGHGDTRLKEDLADAMRESPTSAAARRETTAASVDQAAPPCGDHLATGASPASADLAA	956	
QY	156	GTHSPMGLNFDSELVDSFPD---QNFHEVCSGIHPDEAAEKMTSV-----VAEN	203	
Db	957	GAGGSASSNDNVCYYDALAAHEHSLNDLDEVEDQHTDAETHLSHMSDISALASVLPN	1016	
QY	204	GTGLVCSLELEEXQPELKMCGYNGSPSV-----ESLHQEVSVLVPDPTV---SC	250	
Db	1017	-----ICSYSRVE-----KWNPLGTVDLWAGISGLQTDMLDEVVQGLAAEDTVIHENC	1067	
QY	251	LDDPSHLP-----DQLEDPTILSED-----SLEP-----FNSLAPEPV	283	
Db	1068	-DHALFPVVAIFCSIPAEROKLLQDKDQTEATASELEREQASCQPPVVLVKNKLIVPVI	1126	
QY	284	SGGLYGI-----DDTELMGAEEDKLPLXDSFVIALDCP--	316	

Db 1127 AKEMDEVLVQKQKQKQSHKAAQQQQQQQQQQKTLAAAEADAAQAQKFAELSHLNKPKG 1186
Qy 317 -SLNATAF-----SLLADDSOTSTSIFA-----SPTSPVVLGESVLDNSFDLNN 361
Db 1187 NSNNKNTASQSGSTSEDEATTTITLAKAKAKNNKQKTSGRVSEKKPEQTNL----- 1241
Qy 362 GSDAEQ----BEMETQSSDFPPLTQPADPDOSSTIQLHPATSPAVSPPTTSPAVSLVSPA 417
Db 1242 ---AEEKVAVEKEENEERAVPMVKTTINKEQSDAESHADSLANKSSIAAV--MVSSA 1296
Qy 418 ASPEISPEVCPAASTVSPVSVSPASSAVLPAVSLVPLTASVTSPKASPTVSPA 477
Db 1297 SAQGLSLHVMSAAD-----AEQGEDEIEGLDEEPPKTKMSKNNKKQ--KPGDA 1344
Qy 478 PPTASPANKDVSSPFLTTADVEITGEGLTAGSGDVMRRRIATPEEVRLPLQHGWRREV 537
Db 1345 VATWT- IDKEKEKAKEKLKEREAKLOEKEKEKLKEREESLRM-----ERE 1397
Qy 538 RIKKSHRWQGETWYYPGCGRMKQFPBVIKYLNRNLVHVSRRHFSPSRMPVGDFFEE 597
Db 1398 KLKE-----EKIKEKE--EKIKEEK-----IKE 1420
Qy 598 RDTPEGLQWVLSABEIPSIQAITGKRGRPRNTEKAKTKVVKVGRGRPPPKVITEL 657
Db 1421 KOREEKLKEEKLKEKEREEMK-----EKEREKAKQREKREKREKREKREBKL 1473
Qy 658 LNKTDNRPLKLEAQETLNEE-----DKAKIAKSKQMR-QKVORGECLTTIQOARNK 710
Db 1474 KEKLREKIKIKEKEEKLKEREKEREKEREKEREKEREVEKIKEKEREKREKREKREK 1533
Qy 711 RKQETKSLKHEKAKKSAEKEGKTQKELKVKY-----REKKEKVKMKEK 758
Db 1534 LKEKEELKKEKEKEKEKEEKLKEREKQELKEREKREKREKREKREKREKREKREK 1593
Qy 759 BEVTKAKPACAKDTLATQRLERQKQOMILEEMK----- 795
Db 1594 ERAEKLKOLEKVKLKEKEEQKKEKELKKEKEDKVKKEKSEKELLISATVSN 1653
Qy 796 -----PTEDMCLTHQPL-----PDFSR-----VFGLLPLPSGAFSDCTIVEFL 834
Db 1654 PMRRVVEDTPPKLPAVQDPSLGLKPKTKASPEKRDCKLFLGLTTPPKEVND--TFEDFL 1711
Qy 835 HSFGKVLGFDPAKDVPSLGVLOGLLCOGSLGEGVODLLVRLKAAALHDPGPPSPCOSLK 894
Db 1712 S-----GLKLEALPPLPAL-EPLEVKEDSKKSVSLI-----NFEPLQETA 1753
Qy 895 ILGEKVSEIP--LTRDNVSEILRCFLMAYGVKXPALC-----DLRLTOPFOAQPPOQ 943
Db 1754 SIPRKISPPPRGFTQNL--IL-----ALCGSLHYENEQERIRETEAEVQP--- 1797
Qy 944 KAAVLAFVPHELNGSTLIINEIDKTLISMSS-----YRKNKIVVEGLRLRL 989
Db 1798 EVWTTPEPVY-ITDLTALQKSTSTSNNSGSEIIVVMEEPVKLSKHKRLKQLQLOQQ 1856
Qy 990 KTVLAKTRGSEVEMGREPEELGRRSS-----RIM-EETSMEEEEEESIAA VP- 1039
Db 1857 Q---QOREPDEDELRPLISIMCDOLDVPDIPVRAMPDDDEGVBEETEPETEPEQSUPE 1913
Qy 1040 -----GRRG-----RRDGEVDATAS-----SIPELERQIEKLSKRLQFRFKLLH 1079
Db 1914 DLLHFGSGSVATNATTTDSEGPVPATTTSDNNVVYGSSTSNAAPHKLK-----TKKLEH 1967
Qy 1080 SSQMLRAVSLGODRVRRRYWVLPYLAGIFVSGTEGNLVPVEVIKKTDSLKVAHASLNP 1139
Db 1968 KINLIAATEAA-----AGSNTT-----ASSPARAERSPLKTKPGFMQPRHFKSVYRG 1184
Qy 1140 ALFSMKMEL-----AGSNTT-----ASSPARAERSPLKTKPGFMQPRHFKSVYRG 1184
Db 2000 TGLGIGIGLOSHSGPAAVAGSSTTGGATAASLPAAGGVTSVGLADG-----APT-A 2050
Qy 1185 QDEQFOAQLOPEAQLHVPAPQPOLQLOLQSHKGFLEQSGPLSLGQSQHDLSQSAFLS 1244
Db 2051 HALQOQQQQQPEQ---PLQPPP----- 2070

Qy 1245 WLSOTQSHSSLLSSSVLTDPSSPGKLDPAQSPPEPEPEDEASSPDLAQAFWNISAQMP 1304
Db 2071 -----PDHPPPEEPGPAIFQPNMLKQEEVEQPLDGLLANNNNIN 2112
Qy 1305 CNAAPTPLAVSEDOPTSPQOLASSKPMRPSAANPCSPVQFSTPLAGLAPKRA-GD 1363
Db 2113 NNIPT-----KONASPNP-----NTNTNPNAIPSSNKVOKAEVTDVDPYPIRHSYGD 2160
Qy 1364 PDEM 1367
Db 2161 HGEL 2164
RESULT 45
ABBS5814
ID ABBS5814 standard; protein; 2897 AA.
XX
AC ABBS5814;
XX
XX 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 2334.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US009231.
XX
PR 23-MAR-2000; 2000US-0191637P.
PR 11-JUL-2000; 2000US-00614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
XX
N-PSDB; ABL02617.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
genes from Drosophila and for elucidating cell signaling and cell-cell
interactions.
XX
PS Disclosure; SEQ ID NO 2334; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
capable of detecting 1000 or more genes from Drosophila. The invention is
useful in developmental biology and in elucidating cell signalling and
cell-cell interactions in higher eukaryotes for the development of
insecticides, therapeutics and pharmaceutical drugs. The invention
discloses genomic DNA sequences (AB16176-AB16175), expressed DNA
sequences (AB16176-AB16175) and the encoded proteins (ABBS57737-
ABBS72072). The sequence data for this patent did not form part of the
printed specification, but was obtained in electronic format directly
from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 2897 AA;
Query Match 3.1%; Score 300; DB 4; Length 2897;
Best Local Similarity 19.4%; Pred. No. 5.5e-10;
Matches 326; Conservative 212; Mismatches 627; Indels 514; Gaps 72;
Qy 330 DSQSTSFASPTSPVVLGESVLDNSFDLNGSDAEQEMETQSSDPPPSLTQAPDQOS 389
Db 233 DEDTSSNSCHPKPPP-----PEQKEEEDDEVDGEPEK----- 266
Qy 390 STIQLHPATSPAVSPTTSPAVSLVSPASPEISPEVCPAASTVSPVSVSPASSAV 449

[illegible]

Db	1107	---RRSGRRGPAVHSELPQPKRTGGARDKQKQPEVNAELEKQESDEDEKISTKIKSEAK	1167
Qy	1403	--PVP-----PEMCSGWWIIPDPDEMLDAMLKALHPRGIRGKALKHLKHKRDFLQEV	1452
Db	1164	DDPAPESPENRKKPEKP---IKEEPNEEPKPKVGRGRPKK-----REVDTTNI	1212
Qy	1453	CLRPSADPTPEPQLPAFQEGIMSWSPKEKTYTDTDLAVLQWVLELQK	1503
Db	1213	ETNDSSETPVQRSSRIA--QOKIKEEAERRKQEEVALRTMK--QELKKKKKAKEADPTVL	1268
Qy	1504	SDIQRGWTCPSPDSRTEDLAYCEHLSDSQEDITWRGPGREGIAPQKTTNPLDLAVML	1563
Db	1269	E-----PSGESESESEAE--EAAKNNKKKKCPGKDGWSSDS	1305
Qy	1564	AALQNVKRRYLREPWPHTHEVVLEKALISTPNGAPEGTTTETSYEITPRIRWRQT---	1620
Db	1306	---EQPESEBEEEP--PHYE-----TDPGSP---LFRSDHEFSPSELEDSQV	1348
Qy	1621	LQCRSAHVCLCLGHLERSIAWEKSVNKTCLVCRKGDNDFFLLCD--GQDRCCHIY	1677
Db	1349	PMKRARTV-----RKENADDLBEDAEACQCKGKSDHPHEWILLCTDPTCNKGHYCS	1400
Qy	1678	CHRPKMEAVPEGDMFCTVCLAQO---VEGEFTQKPGFKRGOKRKSQYSLNFSSEGDGR	1733
Db	1401	CLSPVLFIYPEGDWHCPQCOEQOLIAALERQLEQYDTLVAQKQERI-----LAEEQAE	1455
Qy	1734	RRVLLKGRSPAAAGPRYSBERLSFKRRRLSMNHHSDLTFCFIIIMWESHDAAWPFKE	1793
Db	1456	EQLEEAATLAAKDNFKSEKEDEDDR-----DDMAVGKAEKVKRRRGGR-----	1502
Qy	1794	PVNPRIVS-GYRRIKNPMDFTMRERLLRGVTSSEEFADALLVFNDCQTTNEDDSE	1851
Db	1503	-INRAAKRGTRRRRGNESD-SSHRKSLGSGSRSGSGSDSS-----DNSTSPSDSDE	1554
RESULT	46		
ADL22668			
ID	ADL22668	standard; protein; 601 AA.	
XX	ADL22668;		
AC			
XX			
DT	20-MAY-2004	(first entry)	
XX			
DE		Human disease detection and treatment (MDDT) protein - SEQ ID 117.	
XX			
KW		disease detection; MDDT; antiarteriosclerotic; antipsoriatic; cyostatic;	
KW		cell signalling; arteriosclerosis; psoriasis; cancer; human.	
XX			
OS		Homo sapiens.	
XX			
PN	WO2003062379-A2.		
XX			
PD	31-JUL-2003.		
XX			
XX	14-JAN-2003; 2003WO-US001363.		
XX			
PR	17-JAN-2002; 2002US-0349413P.		
PR	17-JAN-2002; 2002US-0349946P.		
XX			
PA	(INCY-) INCYTE GENOMICS INC.		
XX			
PI	Jones AL, Dahl CR, Gietzen D, Chinn J, Dufour GE, Jackson JL;		
PI	Yu JY, Tuason O, Yap PE, Ameshey SR, Dem TC, Liu TF, Gerstin EH;		
PI	Peralta CH, Lewis SA, Chen AJ, Marwaha R, Lan RY, Urashka ME;		
PI	Kristnam SR, Kolluru V, Panesar IS;		
XX			
DR	WPI; 2003-853443/79.		
DR	N-PSDB; ADL22564.		
XX			
XX			
PT	New isolated disease detection and treatment polynucleotide for		
PT	diagnosing or treating conditions associated with cell signaling e.g.		
PT	arteriosclerosis, psoriasis, and cancer.		
XX			

PS Claim 27; SEQ ID NO 117; 411pp; English.
XX
CC The invention relates to a novel isolated disease detection and treatment
CC (MDT) DNA polynucleotide. The polynucleotide of the invention
CC demonstrates antiarteriosclerotic, antiposoriatic and cytostatic
CC activities and may be useful in a composition for detecting the
CC expression of a disease detection and treatment molecule polynucleotide.
CC The molecules of the invention may be utilised to diagnose or treat
CC conditions, diseases or disorders associated with cell signalling, such
CC as arteriosclerosis, psoriasis and cancer. The current sequence is that
CC of a human MDT protein of the invention.
XX
SQ Sequence 601 AA;

Query Match 3.0%; Score 296; DB 7; Length 601;
Best Local Similarity 20.7%; Pred. No. 9.5e-11;
Matches 151; Conservative 82; Mismatches 273; Indels 222; Gaps 27;

QY 1199 QLVPAQPQQLQLOQSHKGFLEQSGPLSLGOSQH--DLSQSAFLSWLSQTSQSHSLL 1256
Db 1 QQH-PRQAPQV-----LQQPRLISVQTMQRCNNWCGAF-----QAHQWRL 42
QY 1257 SSVLTPDSSPGKLDPAQPPEPEPDEAESPDQAFWFNISAQMPCNAAPTPLAVS 1316
Db 43 AQNA--RIFG-----IPRHSGPOYSMMQPHLQROHSN-----PGHAGPFPVVSVH 86
QY 1317 EDQTPSPQQLASSKPMNR-----PSAANP-----CSPVQFSSTPLA-- 1353
Db 87 NTTINPTSTATTANANRGTPSPVTAIEIPSVTNPENLPSLFDIPPIQANVPMHS 146
QY 1354 --GLAPKRAGDPCGEMPSPTGLQPKRRGPPKFFKQMEQRYLTQLTQAPVPEMCSG 1411
Db 147 WYEGAREKTDQNVLEDA-----GSSLDNLSRY-----ISGHLPPQPTST 190
QY 1412 WWWIPDPEMLDAMLKALHPRGIREKALHKLHNRHDFLQVCLRPSADPIPEPQLPAFQ 1471
Db 191 MNPSPGSPALSPGSSGLS-----NSH-----TPVRPPTSSTSGS 224
QY 1472 EGIMSWSPK--EKT---YETDLAVLOWVEELEQRVINSDQIRGWTCPSPDSTREDLAYC 1526
Db 225 RGSCGSGSRTAECTSLSPKQVVKQBPFGTEDEI-----CSFSGGVKQ----- 268
QY 1527 EHLSDSOEDITWRGPRGLAPQKRTTNPLDLAVNRVLAALRQNVKRYLRPLWPTHEV 1586
Db 269 EKTEDGRSACMLSPSSSLTPPLSTNLHSELDALASLENHVK-----IEPA-DMNESC 323
QY 1587 LEKALLSTPNGAPEGTTTTEISYEITPRIRIWRQTLQRCSAAHVCLCLGHLERSIAWEKS 1646
Db 324 KQGLSLLVNCK-----SPIRSLMHSARIGGGDGN--KDD 358
QY 1647 VNKYTCVLCRKGDNDEFLLCGDGRGCHYIYCHRPKMEAVPEGDWFTVCVLAQVGEFT 1706
Db 359 PNEDWCAVCONGGD---LLCCEKCPKVPHLTCHVPTLLSPSGDWICTFC----- 405
QY 1707 QKPGFPKRGQKSGSYLNFSEGGRRRRVLLKRESAPAGPRYSEERLSPKRRRLSMR 1766
Db 406 -----RDIGKPEVEYDQNLQSKKGTQAQ 431
QY 1767 NHHSDLTFCIILMESHDAWPFEPVNPRLVSGYRRIIKNPMDFSTMERLLR---G 1823
Db 432 LSPVDQRKCRLLLYLCHELSIEFQEV-PASIPNYKIILKPMDLSTVKKLQKHSQ 490
QY 1824 GYTSSEEPADALLVDFNQCQTFNE-----DSEVKGAGHIMRRFPESR 1866
Db 491 HYQIPDDFVADVRLIFKNCFERNEMKVVQVYADTQEIINLKADSEVAQAGKAVLYEDK 550
QY 1867 WEEFVQK 1874
Db 551 LTEIYSR 558
RESULT 47
ABB68397

ID ABB68397 standard; protein; 2768 AA.
XX
AC ABB68397;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 31983.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US009231.
XX
PR 23-MAR-2000; 2000US-0191637P.
PR 11-JUL-2000; 2000US-00614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR N-PSDB; ABL12500.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signaling and cell-cell
PT interactions.
XX
PS Disclosure; SEQ ID NO 31983; 2lpp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABBS7737-
CC AB872072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 2768 AA;
Query Match 3.0%; Score 290.5; DB 4; Length 2768;
Best Local Similarity 19.5%; Pred. No. 2.2e-09;
Matches 357; Conservative 237; Mismatches 628; Indels 605; Gaps 88;
QY 154 ANGTHSPMGL-NFDSQBLYSFPDQNF-----EVCSGIHPDEAAEKMTSVVA---- 201
Db 303 AKGHIAIPREDVDLQEHIDD-DDKNKETIPSAELSGEIEEBEEKDKATTVPVQV 361
QY 202 -----ENGTLVCS---LELEKQPELMKMGVNGSVPSVSLHQVSVLV--PDP 246
Db 362 TDEKDFSEFESSTAGIPSDSRIDLPSSTBEK-----ESSTAAAEEDIVKIITPEP 414
QY 247 TVSCLDL---PSHLPDQLEDTPILSDESLSPFNLSAPEVPSGGLYGIDDTLMGAEDKLP 303
Db 415 EGSGEEDVPKPSQIPEK-----EITEDELIKVSTSAFAKAS-----PEEEVVTATTGAP 463
QY 304 LXDSFVIALDCPSLNNTATAFSLIADDSQ-----TSTISFAS 340
Db 464 TEE-----DVKPTTAGTISEEEEGKPTPAEGSGEEEKDVKVTAAPEETDEAK 513
QY 341 PTSPPLVGSVLQ-----DNSTDL-----NNGSDAQEQEMETQSSDFPSSL 381
Db 514 PTSAPVASDEKEQPKPSEGSDELDLKPATTAGTASASESESEQDEGKSTAPTSTV 573
QY 382 --TQPA-PDQSST-----IQLHPATSPAVSPPTTSPAVSLV--VSPAASP-----EISP 424

RESULT 49	
ID ADE62723	
AC ADE62723 standard; protein; 2468 AA.	
XX AC ADE62723;	
XX DT 29-JAN-2004 (first entry)	
XX Human Protein NP_005900, SEQ ID NO 8656.	
XX DE Human; pain; neuronal tissue; gene therapy;	
KW spinal segmental nerve injury; Chronic constriction injury; CCI;	
KW spared nerve injury; SNI; Chung.	
OS Homo sapiens.	
XX WO2003016475-A2.	
PX 27-FEB-2003.	
XX 14-AUG-2002; 2002WO-US025765.	
XX 14-AUG-2001; 2001US-0312147P.	
PR 01-NOV-2001; 2001US-0346382P.	
PR 26-NOV-2001; 2001US-0333347P.	
XX (GEO) GEN HOSPITAL CORP.	
PA (FARB) BAYER AG.	
XX Woolf C, D'urso D, Befort K, Costigan M;	
PI WPI; 2003-268312/26.	
DR GENBANK; NP_005900.	
XX New composition comprising two or more isolated polypeptides, useful for	
PT preparing a medicament for treating pain in an animal.	
XX Claim 1; Page; 1017pp; English.	
XX The invention discloses a composition comprising two or more isolated rat	
CC or human polynucleotides or a polynucleotide which represents a fragment,	
CC derivative or allelic variation of the nucleic acid sequence. Also	
CC claimed are a vector comprising the novel polynucleotide, a host cell	
CC comprising the vector, a method for identifying a nucleotide sequence	
CC which is differentially regulated in an animal subjected to pain and a	
CC kit to perform the method, an array, a method for identifying an agent	
CC that increases or decreases the expression of the polynucleotide sequence	
CC that is differentially expressed in neuronal tissue of a first animal	
CC subjected to pain, a method for identifying a compound which regulates	
CC the expression of a polynucleotide sequence which is differentially	
CC expressed in an animal subjected to pain, a method for identifying a	
CC compound that regulates the activity of one or more of the	
CC polynucleotides, a method for producing a pharmaceutical composition, a	
CC method for identifying a compound or small molecule that regulates the	
CC activity in an animal of one or more of the polypeptides given in the	
CC specification, a method for identifying a compound useful in treating	
CC pain and a pharmaceutical composition comprising the one or more	
CC polypeptides or their antibodies. The polynucleotide or the compound that	
CC modulates its activity is useful for preparing a medicament for treating	
CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction	
CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene	
CC therapy). The sequence presented is a human protein (shown in Table 2 of	
CC the specification) which is differentially expressed during pain. Note:	
CC The sequence data for this patent did not form part of the printed	
CC specification, but was obtained in electronic form directly from wipo at	
CC ftp.wipo.int/pub/published_pct_sequences.	
XX Sequence 2468 AA;	
SQ Query Match 2.9%; Score 288; DB 7; Length 2468;	
Best Local Similarity 19.2%; Pred. No. 2.8e-09;	
Matches 326; Conservative 249; Mismatches 583; Indels 544; Gaps 82;	

Db 1017 EEEA--DEEDKAEDAREEYEPKMEAEYVMAVVDKAAEAGGAEQYGLTTPTKQLGA 1074
Qy 1156 SSPAR--ARSRPLTKPGFMQPRHFKSPVRGDSQPOAQLOPE-----AQLHVP 1203
Db 1075 OSPGRASSIHDETLPG-----GSESEATASDEENREDQPEEFTATSGYTOSTIBS 1127
Qy 1204 AQPQPOLQLQSHKGFLEQEGSLSL--GOSQHDLSQSAFLSMLSQTSQSHSLSSSVL 1261
Db 1128 SEPTB-----MDEMSIPRDVMSDETNNETESPSQEFVNITKYESSLYQBYIS 1175
Qy 1262 TP-----DSSPECK---LDPAPSPQPEPEPEDEABSSPDLQAFWNISAQM 1303
Db 1176 KPADVTPPLNGFSEGSKTATDGKDYNASATISPPSSMEEDKFSRSALRDAYCEVKAAT 1235
Qy 1304 PCNAAPTPLAVSEDPQTPSPQOLASKPMRNPNSAANPCS--PVQFSFTP----- 1351
Db 1236 TLDINKDIS-AVSSSEKVSFSPSLSPSPPS-PLEKTPLGERSVNFSLTPNEIKVSAEAE 1293
Qy 1352 LAGLAP-----KRRAGDPGEMP--QSPTGLGQPKRR----- 1380
Db 1294 VAPVSEPTQEVVEHCASPEDKTLVVVSPSQSVTGSAGHTPYQSP-----DEKSHLPT 1350
Qy 1381 ---CRPPS-----KFFKQMEQRYLTQLTAQVPPPEMCGMWWIPPEMLDAMLKALHPRG 1432
Db 1351 EVIEKPPAVPVVSFEFSDAKDENERASVSPMDEP-----VPDSE--SPIEKVLS-- 1397
Qy 1433 IREKALHKLHNRDPL---QEVCLRPSADIFE-----PRQLAPQF--GIMSWSPK 1480
Db 1398 LRSPLIGESAYESFLSADDKASGRGAESPFEKSGKQSGPDQVSPVSEMTSTSLYQDK 1457
Qy 1481 EKVETDLAVLQWVEELQORVMSDLOIRGWTCSPDSTREDLAYCEH-----LSDSQEDI 1536
Db 1458 QEGKSTDFAPIK--EDFGQEKKTDDVEAM-----SSQPALADDERKLDGVSPTQIDV 1507
Qy 1537 TWRGPGREGL-----APQRKTNPLDLAVMLAALQONVKRRYLREPLWPTHEVLEK 1589
Db 1508 SQFGSFKEDTKMSISEGTVSKSATPVDEGVAE-----DTYSHMEGV 1549
Qy 1590 ALLSTPNCA-----PEGTTTETS 1607
Db 1550 ASVSTASVATSSFPEPTTDDVS 1571

RESULT 50

Id AD562719 standard; protein; 2468 AA.
Xx AC AD562719;
Xx DT 29-JAN-2004 (first entry)
Xx DE Human Protein AAA18904, SEQ ID NO 8652.
Xx KW Human; pain; neuronal tissue; gene therapy;
Xx KW spinal segmental nerve injury; chronic constriction injury; CCI;
Xx KW spared nerve injury; SNI; Chung.
Xx OS Homo sapiens.
Xx PN WO2003016475-A2.
Xx XX 27-FEB-2003.
Xx XX 14-AUG-2002; 2002WO-US025765.
Xx PF 14-AUG-2001; 2001US-0312147P.
Xx PR 01-NOV-2001; 2001US-0346382P.
Xx PR 26-NOV-2001; 2001US-0333347P.
Xx XX (GEO) GEN HOSPITAL CORP.
Xx PA (FARB) BAYER AG.
Xx XX

PI Woolf C, D'urso D, Befort K, Costigan M;
Xx WPI, 2003-268312/26.
Xx DR GENBANK; AAA18904.
Xx PT New composition comprising two or more isolated polypeptides, useful for
Xx PT preparing a medicament for treating pain in an animal.
Xx PS Claim 1; Page; 1017pp; English.
Xx XX

The invention discloses a composition comprising two or more isolated rat
or human polynucleotides or a polynucleotide which represents a fragment,
CC derivative or allelic variation of the nucleic acid sequence. Also
CC claimed are a vector comprising the novel polynucleotide, a host cell
CC comprising the vector, a method for identifying a nucleotide sequence
CC which is differentially regulated in an animal subjected to pain and a
CC kit to perform the method, an array, a method for identifying an agent
CC that increases or decreases the expression of the polynucleotide sequence
CC that is differentially expressed in neuronal tissue of a first animal
CC subjected to pain, a method for identifying a compound which regulates
CC the expression of a polynucleotide sequence which is differentially
CC expressed in an animal subjected to pain, a method for identifying a
CC compound that regulates the activity of one or more of the
CC polynucleotides, a method for producing a pharmaceutical composition, a
CC method for identifying a compound or small molecule that regulates the
CC activity in an animal of one or more of the polypeptides given in the
CC specification, a method for identifying a compound useful in treating
CC pain and a pharmaceutical composition comprising the one or more
CC polypeptides or their antibodies. The polynucleotide or the compound that
CC modulates its activity is useful for preparing a medicament for treating
CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction
CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
CC therapy). The sequence presented is a human protein (shown in table 2 of
CC the specification) which is differentially expressed during pain. Note:
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic form directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
Xx XX

Query Match 2.9%; Score 288; DB 7; Length 2468;
Best Local Similarity 19.2%; Pred. No. 2.8e-09;
Matches 326; Conservative 249; Mismatches 583; Indels 544; Gaps 82;
Qy 164 NFDQELYDPPDQNFEEVCSGHPDEAAEKEMTSVVAENGTLVCSLELEEXQPELXMC 223
Db 156 SFSQNFIEFTDOEIGELLSTTHP--ANKASLTLFCPEGDWKNNSLNDRNLQDFINIK 213
Qy 224 GYNGSV-PSVESLHQEVSVLVDPDPTVSCLDPPSHLPQLEDTPILSDESL--PFNSLA 279
Db 214 LNSASILPEMEGUS-----EFTEYLSE-SVEVPSFFDILE 247
Qy 280 PEPVSGG-----LYGIDDTLM--GAEDK-----LPLXDS 307
Db 248 P-PTSGGFLXSLKPCCVIFPGRGDSALFAVNGENMLINGSERKSCFWKLIHLDRVDS 306
Qy 308 PVISAL---DCPSLNNATAFSLADDSTSTSPASPTSPVLGESVLQDNSFOLNNGSD 364
Db 307 ILLTHIGDNLPGIN-----SMLQRKIABL-----E 332
Qy 365 AEQEMETQSDPF-----PPSLTQAP-----DQSTIO-----LH 395
Db 333 EQSQGGSTTSDMMKNLISPDLGVVFLNVPENLKNPKNPKKSIKSIACFTLYLNKLS 392
Qy 396 PATSP---AVSPPTTSPAV-----SLVVSFA-ASPEISPEVCFASTVVSFAVS 440
Db 393 MKPEPLFRSVGNTIDPVILFQKMGVGLKLEVMYVLPNPKSKEMQYFMQQTGNTKDKAEF- 451
Qy 441 VVSPASSAVLP---AVSLEVPLTASVTSPKASPTVSPAAAAPTASPAKOVSSFLETTAD 497
Db 452 -----ILPNGQEVLDLPISYLTSSS-----LIVMHPANPAEKIIRVLPFGNST 494
Qy 498 VEITGBGLTAGSGDVMRRRIATPEVRPLQHGWRREVRIKKGSHRWQGTWYVYPCG 557

Db	1398	LKSPPIUGISAYESFLSADDDKASGRGASPPFEKSGKGQSPDQVSPVSEMTTSTLYQDK	1457
Qy	1481	EKTETDTLAVLOWVELEQRVIMSDIQRGTWCPSDPSTREDLAYCEH-----LSDSQEDI	1536
Db	1458	QEKGSTDFAPIK--EDFGGEKKTDDEAM-----SSQPALALDERKLGDVSPQTIDV	1507
Qy	1537	TWRGPCREGI-----APQRKTTNPDLDAVMRLAALQNVRRYLRPLWPHTHEVVLEK	1589
Db	1508	SQGSFKEGTKMSISEGTVSKSATPVDEGVAE-----DYSHMEGV	1549
Qy	1590	ALLSTPNGA-----PEGTTTTIS	1607
Db	1550	ASVSTATSVATSGFPPEPTTDDVS	1571

Search completed: March 9, 2005, 14:05:19
Job time : 191.112 secs

THIS PAGE BLANK (USPTO)